

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 January 2002 (10.01.2002)

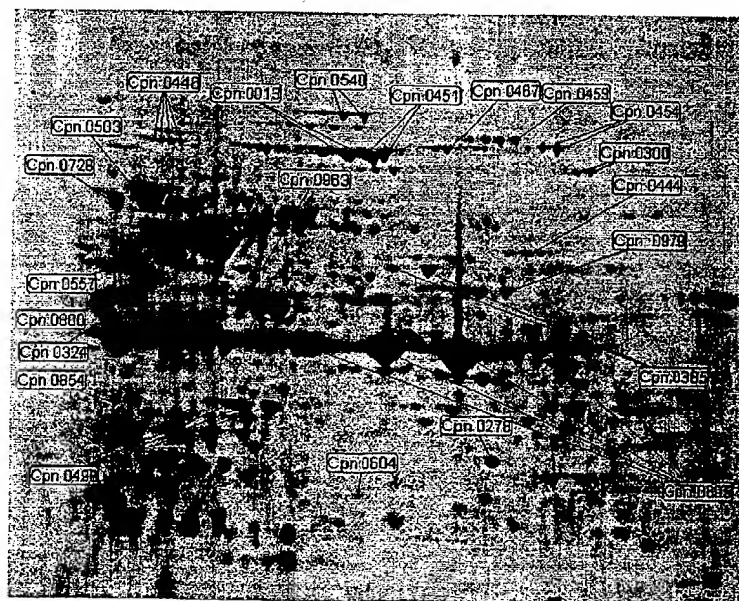
PCT

(10) International Publication Number
WO 02/02606 A2

- (51) International Patent Classification⁷: C07K 14/295, C12N 15/31, A61K 39/118
- (21) International Application Number: PCT/IB01/01445
- (22) International Filing Date: 3 July 2001 (03.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|-----------|--------------------------------|----|
| 0016363.4 | 3 July 2000 (03.07.2000) | GB |
| 0017047.2 | 11 July 2000 (11.07.2000) | GB |
| 0017983.8 | 21 July 2000 (21.07.2000) | GB |
| 0019368.0 | 7 August 2000 (07.08.2000) | GB |
| 0020440.4 | 18 August 2000 (18.08.2000) | GB |
| 0022583.9 | 14 September 2000 (14.09.2000) | GB |
| 0027549.5 | 10 November 2000 (10.11.2000) | GB |
| 0031706.5 | 22 December 2000 (22.12.2000) | GB |
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): RATTI, Giulio [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT). GRANDI, Guido [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT).
- (74) Agents: HALLYBONE, Huw, George et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*



(57) Abstract: The published genomic of *Chlamydia pneumoniae* reveals over 1000 putative encoded proteins but does not itself indicate which of these might be useful antigens for immunisation and vaccination or for diagnosis. This difficulty is addressed by the invention, which provides a number of *C. pneumoniae* protein sequences suitable for vaccine production and development and/or for diagnostic purposes.

WO 02/02606 A2



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention is in the field of immunisation against chlamydial infection, in particular against
5 infection by *Chlamydia pneumoniae*.

BACKGROUND ART

Chlamydiae are obligate intracellular parasites of eukaryotic cells which are responsible for endemic sexually transmitted infections and various other disease syndromes. They occupy an exclusive eubacterial phylogenic branch, having no close relationship to any other known organisms – they are
10 classified in their own order (*Chlamydiales*) which contains a single family (*Chlamydiaceae*) which in turn contains a single genus (*Chlamydia*). A particular characteristic of the *Chlamydiae* is their unique life cycle, in which the bacterium alternates between two morphologically distinct forms, an extracellular infective form (elementary bodies, EB) and an intracellular non-infective form (reticulate bodies, RB). The life cycle is completed with the re-organization of RB into EB, which
15 subsequently leave the disrupted host cell ready to infect further cells.

Four chlamydial species are currently known – *C.trachomatis*, *C.pneumoniae*, *C.pecorum* and *C.psittaci* [e.g. Raulston (1995) *Mol Microbiol* 15:607-616; Everett (2000) *Vet Microbiol* 75:109-126]. *C.pneumoniae* is closely related to *C.trachomatis*, as the whole genome comparison of at least two isolates from each species has shown [Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read
20 *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Stephens *et al.* (1998) *Science* 282:754-759]. Based on surface reaction with patient immune sera, the current view is that only one serotype of *C.pneumoniae* exists world-wide.

C.pneumoniae is a common cause of human respiratory disease. It was first isolated from the conjunctiva of a child in Taiwan in 1965, and was established as a major respiratory pathogen in
25 1983. In the USA, *C.pneumoniae* causes approximately 10% of community-acquired pneumonia and 5% of pharyngitis, bronchitis, and sinusitis.

More recently, the spectrum of *C.pneumoniae* infections has been extended to include atherosclerosis, coronary heart disease, carotid artery stenosis, myocardial infarction, cerebrovascular disease, aortic aneurysm, claudication, and stroke. The association of *C.pneumoniae* with
30 atherosclerosis is corroborated by the presence of the organism in atherosclerotic lesions throughout the arterial tree and the near absence of the organism in healthy arterial tissue. *C.pneumoniae* has also been isolated from coronary and carotid atheromatous plaques. The bacterium has also been associated with other acute and chronic respiratory diseases (e.g. otitis media, chronic obstructive pulmonary disease, pulmonary exacerbation of cystic fibrosis) as a result of sero-epidemiologic
35 observations, case reports, isolation or direct detection of the organism in specimens, and successful

response to anti-chlamydial antibiotics. To determine whether chronic infection plays a role in initiation or progression of disease, intervention studies in humans have been initiated, and animal models of *C.pneumoniae* infection have been developed.

- 5 Considerable knowledge of the epidemiology of *C.pneumoniae* infection has been derived from serologic studies using the *C.pneumoniae*-specific microimmunofluorescence test. Infection is ubiquitous, and it is estimated that virtually everyone is infected at some point in life, with common re-infection. Antibodies against *C.pneumoniae* are rare in children under the age of 5, except in developing and tropical countries. Antibody prevalence increases rapidly at ages 5 to 14, reaching 50% at the age of 20, and continuing to increase slowly to ~80% by age 70.
- 10 A current hypothesis is that *C.pneumoniae* can persist in an asymptomatic low-grade infection in very large sections of the human population. When this condition occurs, it is believed that the presence of *C.pneumoniae*, and/or the effects of the host reaction to the bacterium, can cause or help progress of cardiovascular illness.

- 15 It is not yet clear whether *C.pneumoniae* is actually a causative agent of cardiovascular disease, or whether it is just artefactually associated with it. It has been shown, however, that *C.pneumoniae* infection can induce LDL oxidation by human monocytes [Kalayoglu *et al.* (1999) *J. Infect. Dis.* 180:780-90; Kalayoglu *et al.* (1999) *Am. Heart J.* 138:S488-490]. As LDL oxidation products are highly atherogenic, this observation provides a possible mechanism whereby *C.pneumoniae* may cause atheromatous degeneration. If a causative effect is confirmed, vaccination (prophylactic and
- 20 therapeutic) will be universally recommended.

- Genomic sequence information has been published for *C.pneumoniae* [Kalman *et al.* (1999) *supra*; Read *et al.* (2000) *supra*; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994] and is available from GenBank. Sequencing efforts have not, however, focused on vaccination, and the availability of genomic sequence does not in itself indicate which of the >1000
- 25 genes might encode useful antigens for immunisation and vaccination. WO99/27105, for instance, implies that every one of the 1296 ORFs identified in the *C.pneumoniae* strain CM1 genome is a useful vaccine antigen.

- It is thus an object of the present invention to identify antigens useful for vaccine production and development from amongst the many proteins present in *C.pneumoniae*. It is a further object to
- 30 identify antigens useful for diagnosis (e.g. immunodiagnosis) of *C.pneumoniae*.

DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *C.pneumoniae* amino acid sequences disclosed in the examples.

- It also provides proteins comprising sequences which share at least x% sequence identity with the
- 35 *C.pneumoniae* amino acid sequences disclosed in the examples. Depending on the particular

sequence, x is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more). These include mutants and allelic variants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH
5 program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the *C.pneumoniae* amino acid sequences disclosed in the examples. The fragments should comprise at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (e.g. 8, 10, 12,
10 14, 16, 18, 20, 30, 40, 50, 75, 100 or more). Preferably the fragments comprise one or more epitope(s) from the sequence. Other preferred fragments omit a signal peptide.

The proteins of the invention can, of course, be prepared by various means (e.g. native expression, recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (e.g. native, fusions *etc.*). They are preferably prepared in substantially pure form (*ie.* substantially
15 free from other *C.pneumoniae* or host cell proteins). Heterologous expression in *E.coli* is a preferred preparative route.

According to a further aspect, the invention provides nucleic acid comprising the *C.pneumoniae* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences which share at least $x\%$ sequence identity with the *C.pneumoniae* nucleotide
20 sequences disclosed in the examples. Depending on the particular sequence, x is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the *C.pneumoniae* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (e.g. 65°C in a 0.1xSSC, 0.5% SDS solution).

25 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least n consecutive nucleotides from the *C.pneumoniae* sequences and, depending on the particular sequence, n is 10 or more (e.g. 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 75, 100, 200, 300 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

30 It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (e.g. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (e.g. by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (e.g. single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed therewith.

- 5 According to a further aspect, the invention provides immunogenic compositions comprising protein and/or nucleic acid according to the invention. These compositions are suitable for immunisation and vaccination purposes. Vaccines of the invention may be prophylactic or therapeutic, and will typically comprise an antigen which can induce antibodies capable of inhibiting (a) chlamydial adhesion, (b) chlamydial entry, and/or (c) successful replication within the host cell. The vaccines
10 preferably induce any cell-mediated T-cell responses which are necessary for chlamydial clearance from the host.

- The invention also provides nucleic acid or protein according to the invention for use as medicaments (*e.g.* as vaccines). It also provides the use of nucleic acid or protein according to the invention in the manufacture of a medicament (*e.g.* a vaccine or an immunogenic composition) for
15 treating or preventing infection due to *C.pneumoniae*.

The invention also provides a method of treating (*e.g.* immunising) a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid or protein according to the invention.

According to further aspects, the invention provides various processes.

- 20 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

- A process for detecting *C.pneumoniae* in a sample is provided, wherein the sample is contacted with
25 an antibody which binds to a protein of the invention.

A summary of standard techniques and procedures which may be employed in order to perform the invention (*e.g.* to utilise the disclosed sequences for immunisation) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

- 30 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *e.g.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989) and *Third Edition* (2001); *DNA Cloning, Volumes I and II* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I.
35

- Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

Definitions

- 10 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" e.g. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- 15 The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Chlamydial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been
20 assembled in a single protein in an arrangement not found in nature.

- An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be
25 reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

- A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence
30 identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination,
35 has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (e.g. see US patent 5,753,235).

Expression systems

The Chlamydial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

- 5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*e.g.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA
10 synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

- 15 Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive
20 cells.

- The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription
25 initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al.
30 (1982) *PNAS USA* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

- A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be
35 cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader

fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

- 5 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

- 15 Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

- 30 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion, electroporation, encapsulation of polynucleotide(s) in liposomes, direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g. Hep G2), and a number of other cell lines.

35 ii. Baculovirus Systems

- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence
- 40

homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

5 After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

10 Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its
15 own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

20 Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

25 The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

30 Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*e.g.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

35 Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

40 DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals

for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

10 A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

15 Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

20 After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For
25 example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

30 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between ~1% and ~5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels
35 in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the
40 art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus)

or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, e.g. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, e.g. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, e.g. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

25 iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillan, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilkink and Dons, 1993, *Plant Mol. Biol. Reprtr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's spliceosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be

transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*e.g.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g-laotamase* (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological*

Regulation and Development: Gene Expression (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

5 A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (BPO-A-0 219 237).

10 Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be
15 made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign
20 protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *BioTechnology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is
25 either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva
30 *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct
35 the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See e.g. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic*

- Engineering: *Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem.* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus],
 5 [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, Streptococcus].

v. Yeast Expression

- 10 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may
 15 also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.
- 20 Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].
- 25 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*,
 30 *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119;
 35 Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always

be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

5 Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *e.g.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin
10 region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*e.g.* WO 88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion
15 in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the genes for invertase (EP-A-0012873; JPO 62,096,086) and A-factor (US patent 4,588,684). Alternatively, leaders of non-yeast origin exist, such as an interferon leader, that also provide for secretion in yeast (EP-A-0060057).
20

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion
25 include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*e.g.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct
30 the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression
35 constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YE_p24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy
40

number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *e.g.* Brake *et al.*, *supra*.

- 5 Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be
10 directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the
15 expression construct in the vector, which can result in the stable integration of only the expression construct.

- Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable
20 marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

- Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or
25 developed into an integrating vector, as described above.

- Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].
30
35

- Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *e.g.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze
40 *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459;

- Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patents 4,837,148 & 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach & Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

Pharmaceutical Compositions

Pharmaceutical compositions can comprise polypeptides and/or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

- 5 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

- 10 Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

- Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.
- 15
20

- Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in Vaccine design: the subunit and adjuvant approach, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.
- 25
30
35

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

5 The immunogenic compositions (e.g. the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

10 Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

15 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be
20 determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, e.g. by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (e.g. WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be
25 administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [e.g. Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

30 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

35 The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses e.g. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

- 10 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.
- 15 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (e.g. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.
- 20 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from
- 25 depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671,

WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include

5 adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the

10 remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1995) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is

15 SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are

25 herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 & WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those

30 deposited with ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in

35 US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

40 DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems.

Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

5 Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

10 Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585.

15 Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

20 Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

25 A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

30 Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for recombinant protein expression. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

35 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in e.g. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

5 Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

10 One example are polypeptides which include, without limitation: asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

15 B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

20 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

25 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

30 Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy]propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See,

also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See e.g. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, & E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, & E apoproteins, LDL comprises apoprotein B; HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in

conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

- Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

10 F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

- Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

- The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

- The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Nucleic Acid Hybridisation

- "Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] vol.2, chapt.9, pp.9.47 to 9.57.

“Stringency” refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA
5 immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to
10 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with
15 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length
20 and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4\%[(G + C)] - 0.6(\% \text{ formamide}) - 600/n - 1.5(\% \text{ mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

25 In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in
30 gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with
35 is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and

reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

5 Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

10 The nucleic acid probes will hybridize to the Chlamydial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Chlamydial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

15 The probe sequence need not be identical to the Chlamydial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Chlamydial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Chlamydial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe
20 sequence has sufficient complementarity with the a Chlamydial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more
25 preferably ≥ 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

30 The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *e.g.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*e.g.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*e.g.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].
35

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 & 4,683,202. Two 'primers' hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its

complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Chlamydial sequence.

5 A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Chlamydial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [supra]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The
10 solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-189 show data pertaining to examples 1-189.

Figure 190 shows a representative 2D gel of proteins in elementary bodies.

15 Figure 191 shows an alignment of sequences in five (six) proteins of the invention.

EXAMPLES

The examples indicate *C.pneumoniae* proteins, together with evidence to support the view that the proteins are useful antigens for vaccine production and development or for diagnostic purposes. This evidence takes the form of:

- 20 • Computer prediction based on sequence information from CWL029 strain (e.g. using the PSORT algorithm available from www.psort.nibb.ac.jp).
- Data on recombinant expression and purification of the proteins cloned from IOL207 strain.
- Western blots to demonstrate immunoreactivity in serum (typically a blot of an EB extract of *C.pneumoniae* strain FB/96 stained with mouse antiserum against the recombinant protein).
- 25 • FACS analysis of *C.pneumoniae* bacteria or purified EBs to confirm accessibility of the antigen to the immune system (see also table III).
- An indication if the protein was identified by MALDI-TOF from a 2D gel electrophoresis map of proteins from purified elementary bodies from strain FB/96. This confirms that the protein is expressed *in vivo* (see also table V).
- 30 Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

- ### CLONING OF CPN ORFs FOR EXPRESSION IN *E. COLI*

10 a) proteins having an hexa-histidine tag at the C-terminus (cpn-His)
 b) proteins having a GST fusion partner at the N-terminus (Gst-cpn)
 c) proteins having both hexa-histidine tag at the C-terminus and GST at the N-terminus
 (GST/His fusion; NH₂-GST-cpn-(His)₆-COOH)

20 The modified versions of pGEX-KG vector were made with the aim of allowing the cloning of single amplification products in all three vectors after only one double restriction enzyme digestion and to minimise the presence of extraneous amino acids in the final recombinant proteins.

Two couples of complementary oligodeoxyribonucleotides were synthesised using the DNA synthesiser ABI394 (Perkin Elmer) and the reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) were annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50µl and then were left to cool slowly at room temperature. With the described procedure the following DNA linkers were obtained:

30 NdeI NheI XmaI EcoRI NcoI SalI XhoI SacI NotI
GATCCCATATGCTAGCCCGGGGAATTCGTCCATGGAGTCGAGCTGCAGCTGCAGTGCAGCTCCTGAGCGGGCCGATGAA
GGTATACCGATCGGGCCCTTAAGCAGGTACCTCATCAGCTGACTGAGCTCACTAGCTCGAGGACTCGCCGGCGTACTTTCGA

35 HindIII NotI XhoI --Hexa-Histidine--
TCGACAAGCTTGGCGGCCGACTCGAGCATCACCATCACCATCACTGAT
GTTCGAACCCGGCGTGAGCAGTAGAGGTAGTGGTAGTGACTATCGA

BNSDOCID: <WO__0202606A2_I_>

(New england Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

- The new plasmid pGEX-NN was digested with SalI and HindIII and ligated to the linker gexNNH.
- 5 After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

(B) Chromosomal DNA preparation

- The chromosomal DNA of elementary bodies (EB) of *C.pneumoniae* strain 10L-207 was prepared by
- 10 adding 1.5 ml of lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA, 0,6 % SDS, 100 µg/ml Proteinase K, pH 8) to 450 µl EB suspension (400.000/µl) and incubating overnight at 37 °C. After sequential extraction with phenol, phenol-chloroform, and chloroform, the DNA was precipitated with 0,3 M sodium acetate, pH 5,2 and 2 volumes of absolute ethanol. The DNA pellet was washed with 70 % ethanol. After solubilization with distilled water and treatment with 20 µg/ml RNase A
- 15 for 1 hour at RT, the DNA was extracted again with phenol-chloroform, alcohol precipitated and suspended with 300 µl 1 mM Tris-HCl pH 8,5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample.

(C) Oligonucleotide design

- Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF
- 20 using the sequence of *C.pneumoniae* strain CWL029. Any predicted signal peptide were omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most ORFs, the 5' tail of the primers (table I) included only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (tableI) included a XhoI or a NotI or a HindIII restriction site.

5' tails		3' tails	
NdeI	5' GTGCGTCATATG 3'	XhoI	5' GCGTCTCGAG 3'
NheI	5' GTGCGTGCTAGC 3'	NotI	5' ACTCGCTAGCGGCCGC 3'
SpeI	5' GTGCGTACTAGT 3'	HindIII	5' GCGTAAGCTT 3'

25 **Table I.** Oligonucleotide tails of the primers used to amplify Cpn genes.

- As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the primers which was determined as described [(Breslauer *et al.* (1986) *PNAS USA* 83:3746-50)]. The average melting temperature of the selected oligos was 50-55°C
- 30 for the hybridizing region alone and 65-75°C for the whole oligos. Table II shows the forward and reverse primers used for each amplification.

(D) Amplification

The standard PCR protocol was as follow: 50 ng genomic DNA were used as template in the presence of 0,2 μ M each primer, 200 μ M each dNTP, 1,5 mM $MgCl_2$, 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100 μ l. Each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridizing temperature the one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole lenght primers. The standard cycles were as follow:

10	denaturation : 94 °C, 2 min	
	denaturation: 94 °C, 30 seconds	} 5 cycles
	hybridization: 51 °C, 50 seconds	
	elongation: 72 °C, 1 min or 2 min and 40 sec	
15	denaturation: 94 °C, 30 seconds	} 25 cycles
	hybridization: 70 °C, 50 seconds	
	elongation: 72 °C, 1 min or 2 min and 40 sec	
	72 °C, 7 min	
20	4 °C	

The elongation time was 1 min for ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications were performed using a Gene Amp PCR system 9600 (Perkin Elmer).

25 To check the amplification results, 4 μ l of each PCR product was loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products were loaded on agarose gel and after electrophoresis the right size bands were excised from the gel. The DNA was purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA was 30 50 μ l TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One μ l of each purified DNA was loaded onto agarose gel to evaluate the yield.

(E) Digestion of PCR fragments

One-two μ g of purified PCR product were double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100 μ l final 35 volume. The restriction enzymes and the digestion buffers were from New England Biolabs. After

purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl was subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

(F) Digestion of the cloning vectors (pET21b+, pGEX-NN, and pGEX-NNH)

- 5 10 µg of plasmid was double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample.

10 **(G) Cloning**

75ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each ORF, were ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions were incubated overnight at 16 °C.

- 15 Transformation in *E. coli* DH5 competent cells was performed as follow: the ligation reaction was mixed with 200 µl of competent DH5 cells and incubated on ice for 30 min and then at 42 °C for 90 seconds. After cooling on ice, 0.8 ml LB was added and the cells were incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions were plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates were incubated overnight at 37 °C. The screening of the
20 transformants was done by growing randomly chosen clones in 6 ml LB 100 µg/ml Ampicillin, by extracting the DNA using the Qiagen Qiaprep Spin Miniprep Kit following the manufacturer instructions, and by digesting 2 µl of plasmid miniprep with the restriction enzymes specific for the restriction cloning sites. After agarose gel electrophoresis of the digested plasmid mini-preparations, positive clones were chosen on the basis of the correct size of the restriction fragments,
25 as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

(H) Expression

- 1 µl of each right plasmid mini-preparation was transformed in 200 µl of competent *E. coli* strain suitable for expression of the recombinant protein. All pET21b+ recombinant plasmids were
30 transformed in BL21 DE3 (Novagen) *E. coli* cells, whilst all pGEX-NN and all pGEX-NNH recombinant plasmids were transformed in BL21 cells (Novagen). After plating transformation mixtures on LB/Amp agar plates and incubation overnight at 37 °C, single colonies were inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture was inoculated in 2 ml LB/Amp and grown at 37 °C until OD₆₀₀ of the pET clones reached the 0,4-0,8
35 value or until OD₆₀₀ of the pGEX clones reached the 0,8-1 value. Protein expression was then

induced by adding IPTG (Isopropil β -D thio-galacto-piranoside) to the mini-cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 3 hours incubation at 37 °C the final OD₆₀₀ was checked and the cultures were cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet was suspended in 50 μ l of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerol, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD₆₀₀ culture was analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

PURIFICATION OF THE RECOMBINANT PROTEINS

Single colonies were inoculated in 25 ml LB 100 μ g/ml Ampicillin and grown at 37 °C overnight. The overnight culture was inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD₆₀₀ 0.4-0.8 value for the pET clones, or until OD₆₀₀ 0.8-1 value for the pGEX clones. Protein expression was then induced by adding IPTG to the cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 4 hours incubation at 25 °C the final OD₆₀₀ was checked and the cultures were cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet was processed for purification or frozen at -20 °C.

(I) Procedure for the purification of soluble His-tagged proteins from *E.coli*

1. Transfer the pellets from -20°C to ice bath and reconstitute with 10 ml 50 mM NaHPO₄ buffer, 300 mM NaCl, pH 8.0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline:
2. Break the pellets in the French Press performing three passages with in-line washing.
3. Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)
4. Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8.0.
5. Store the centrifugation pellet at -20°C, and load the supernatant in the columns.
6. Collect the flow through.
7. Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8.0.
8. Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0.
9. Elute the proteins bound to the columns with 4.5 ml (1.5 ml + 1.5 ml + 1.5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0 and collect the 3 corresponding fractions of ~1.5 ml each. Add to each tube 15 μ l DTT 200 mM (final concentration 2 mM)

10. Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 µg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 µl + 7 µl loading buffer).
11. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
- 5 12. For immunisation prepare 4-5 aliquots of 100 µg each in 0,5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at -20°C until immunisation.

(J) Purification of His-tagged proteins from Inclusion bodies

Purifications were carried out essentially according the following protocol:

- 10 1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8,5 on an ice bath.
2. Disrupt the resuspended bacteria with a French Press, performing two passages.
3. Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).
- 15 4. Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8,5. Stir for ~ 10 min. with a magnetic bar.
5. Centrifuge as described above, and collect the supernatant..
- 20 6. Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nickel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H₂O and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8,5.
7. Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-HCl buffer, 1 mM TCEP, 6M urea, pH 8,5
- 25 8. Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8,5. Collect and set aside the first 5 ml for possible further controls.
9. Elute the proteins bound to the columns with 4,5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8,5. Add the elution buffer in three 1,5 ml aliquots, and collect the corresponding 3 fractions. Add to each fraction 15 µl DTT (final concentration 2 mM) .
- 30 10. Measure eluted protein concentration with the Bradford method, and analyze aliquots of ca 10 µg of protein by SDS-PAGE.
11. Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0,5 M arginine, 2 mM DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8,5
- 35

(K) Procedure for the purification of GST-fusion proteins from *E.coli*

1. Transfer the bacterial pellets from -20°C to an ice bath and resuspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (CØMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added. Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:
 - a) Position the probe at about 0,5 cm from the bottom of the tube
 - b) Block the tube with the clamp
 - c) Dip the tube in an ice bath
 - d) Set the sonicator as follows: Timer \rightarrow Hold, Duty Cycle \rightarrow 55, Out. Control \rightarrow 6.
 - e) perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses + ~45" hold; b. 10 impulses + ~45" hold; c. 10 impulses + ~45" hold; d. 10 impulses + ~45" hold; e. 10 impulses + ~45" hold)
2. Centrifuge at about $30-40000 \times g$ for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.
3. Store the centrifugation pellets at -20°C , and load the supernatants on the chromatography columns, as follows
4. Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml ($\cong 1$ ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1) H_2O , and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.
5. Load the supernatants on the columns and discard the flow through.
6. Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7,4.
7. Elute the proteins bound to the columns with 4,5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8,0, adding 1,5 ml + 1,5 ml + 1,5 ml and collecting the respective 3 fractions of ~1,5 ml each.
8. Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10 μg aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21 μl (+ 7 μl loading buffer).
9. Store the collected fractions at $+4^{\circ}\text{C}$ while waiting for the results of the SDS-PAGE analysis.
10. For each protein destined to the immunisation prepare 4-5 aliquots of 100 μg each in 0,5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8,0. Store the aliquots at -20°C until immunisation..

SEROLOGY**(L) Protocol of immunization**

1. Groups of four CD1 female mice aged between 6 and 7 weeks were immunized with 20 μg of recombinant protein resuspended in 100 μl .

2. Four mice for each group received 3 doses with a 14 days interval schedule.
3. Immunization was performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses.
- 5 4. Sera were collected before each immunization. Mice were sacrificed 14 days after the third immunization and the collected sera were pooled and stored at -20°C .

(M) Western blot analysis of Cpn elementary body proteins with mouse sera

- Aliquots of elementary bodies containing approximately 4 μg of proteins, mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at 95°C , were loaded on a 12% SDS-PAGE gel. The gel was run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel was electroblotted onto nitrocellulose membrane at 200 mA for 30 minutes. The membrane was blocked for 30 minutes with PBS, 3% skimmed milk powder and incubated O/N at 4°C with the appropriate dilution (1/100) of the sera. After washing twice with PBS + 0.1% Tween (Sigma) the
- 15 membrane was incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Sigma) diluted 1:3000. The nitrocellulose was washed twice for 10 minutes with PBS + 0.1% Tween-20 and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

Lanes shown in Western blots are: (P) = pre-immune control serum; (I) = immune serum.

(N) FACS analysis of *Chlamydia pneumoniae* elementary bodies with mouse sera

- 20 1. 2×10^5 Elementary Bodies (EB)/well were washed with 200 μl of PBS-0.1%BSA in a 96 wells U bottom plate and centrifuged for 10 min. at 1200rpm, at 4°C .
2. The supernatant was discarded and the E.B. resuspended in 10 μl of PBS-0.1%BSA.
3. 10 μl mouse sera diluted in PBS-0.1%BSA were added to the E.B. suspension to a final dilution of 1:400, and incubated on ice for 30 min.
- 25 4. EB were washed by adding 180 μl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C .
5. The supernatant was discarded and the E.B. resuspended in 10 l of PBS-0.1%BSA.
6. 10 μl of a goat anti-mouse IgG, F(ab')₂ fragment specific-R-Phycoerythrin-conjugated (Jackson Immunoresearch Laboratories Inc., cat.N^o115-116-072) was added to the EB suspension to a final dilution of 1:100, and incubated on ice for 30 min. in the dark.
- 30 7. EB were washed by adding 180 μl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C .
8. The supernatant was discarded and the E.B. resuspended in 150 μl of PBS-0.1%BSA.
9. E.B. suspension was passed through a cytometric chamber of a FACS Calibur (Becton Dickinson, Mountain View, CA USA) and 10.000 events were acquired.

10. Data were analysed using Cell Quest Software (Becton Dickinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on E.B. signals. An histogram plot was then created on FL2 intensity of fluorescence log scale recalling the morphological region of EB.
- 5 NB: the results of FACS depend not only on the extent of accessibility of the native antigens but also on the quality of the antibodies elicited by the recombinant antigens, which may have structures with a variable degree of correct folding as compared with the native protein structures. Therefore, even if a FACS assay appears negative this does not necessarily mean that the protein is not abundant or accessible on the surface. PorB antigen, for instance, gave negative results in FACS but is a surface-
- 10 exposed neutralising antigen [Kubo & Stephens (2000) *Mol. Microbiol.* 38:772-780].

(O) Mass Spectrometry analysis of two-dimensional electrophoretic protein maps

- Gradient purified EBs from strain FB/96 were solubilized at a final concentration of 5.5mg/ml with immobiline rehydration buffer (7M urea, 2M thiourea, 2% (w/v) CHAPS, 2% (w/v) ASB 14 [Chevallet *et al.* (1998) *Electrophor.* 19:1901-9], 2% (v/v) C.A 3-10NL (Amersham Pharmacia
- 15 Biotech), 2 mM tributyl phosphine, 65 mM DTT). Samples (250µg protein) were adsorbed overnight on Immobiline DryStrips (7 cm, pH 3-10 non linear). Electrophocusing was performed in a IPGphor Isoelectric Focusing Unit (Amersham Pharmacia Biotech). Before PAGE separation, the focused strips were incubated in 4M urea, 2M thiourea, 30% (v/v) glycerol, 2% (w/v) SDS, 5mM tributyl phosphine 2.5%(w/v) acrylamide, 50mM Tris-HCl pH 8.8, as described [Herbert *et al.* (1998)
- 20 *Electrophor.* 19:845-51]. SDS-PAGE was performed on linear 9-16% acrylamide gradients. Gels were stained with colloidal Coomassie (Novex, San Diego) [Doherty *et al.* (1998) *Electrophor.* 19:355-63]. Stained gels were scanned with a Personal Densitometer SI (Molecular Dynamics) at 8 bits and 50µm per pixel. Map images were annotated with the software Image Master 2D Elite, version 3.10 (Amersham Pharmacia Biotech). Protein spots were excised from the gel, using an Ettan
- 25 Spot picker (Amersham Pharmacia Biotech), and dried in a vacuum centrifuge. In-gel digestion of samples for mass spectrometry and extraction of peptides were performed as described by Wilm *et al.* [*Nature* (1996) 379:466-9]. Samples were desalted with a ZIP TIP (Millipore), eluted with a saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA and directly loaded onto a SCOUT 381 multiprobe plate (Bruker). Spectra were acquired on a Bruker
- 30 Biflex II MALDI-TOF. Spectra were calibrated using a combination of known standard peptides, located in spots adjacent to the samples. Resulting values for monoisotopic peaks were used for database searches using the computer program Mascot (www.matrixscience.com). All searches were performed using an error of 200-500ppm as constraint. A representative gel is shown in Figure 190.

Example 1

- 35 The following *C.pneumoniae* protein (PID 4376552) was expressed <SEQ ID 1; cp6552>:

1 MKKKLSLLVG LIFVLSSCHK EDAQNKIRIV ASPTPHAELL ESLQEEAKDL

51 GIKLKILPVD DYRIPNRLLL DKQVDANYFQ HQAFLDDECE RYDCKGELVW
 101 IAKVHLEPQA IYSKKHSSLE RLKSQKKLTI AIPVDRTNAQ RALHLLBECG
 151 LIVCKGPANL NMTAKDVCCK ENRSINILEV SAPLLVGSPL DVDAAVIPGN
 201 FAIAANLSPK KDSLCLLEDLS VSKYTNLVVI RSEDVGSPEM IKLQKLFQSP
 251 SVQHFFDTKY HGNILTMTOQ NG*

A predicted signal peptide is highlighted.

The cp6552 nucleotide sequence <SEQ ID 2> is:

1 ATGAAAAAAA AATTATCATT ACTTGTAGGT TTAATTTTGT TTTGAGTTC
 51 TTGCCATAAG GAAGATGCTC AGAATAAAAT ACGTATTGTA GCCAGTCCGA
 101 CACCTCATGC GGAATTATTG GAGAGTTTAC AGGAAGAGGC TAAAGATCTT
 151 GGAATCAAGC TGAAAATACT TCCAGTAGAT GATTATCGTA TTCCTAATCG
 201 TTTGCTTTTG GATAACAAG TAGATGCAA TTACTTTCAA CATCAAGCTT
 251 TTCTTGATGA CGAATGCGAG CGTTATGATT GTAAGGGTGA ATTAGTTGTT
 301 ATCGCTAAAG TTCATTTGGA ACCTCAAGCA ATTTATTCTA AGAAACATTC
 15 351 TTCTTTAGAG CGCTTAAAAA GCCAGAAGAA ACTGACTATA GCGATTCCCTG
 401 TGGATCGTAC GAATGCTCAG CGTGCTCTAC ACTTGTTAGA AGAGTGCGGA
 451 CTCATTGTTT GCAAAGGGCC TGCTAATTTA AATATGACAG CTAAAGATGT
 501 CTGTGGGAAA GAAAATAGAA GTATCAACAT ATTAGAGGTG TCAGCTCCTC
 551 TTCTTGTCGG ATCTCTTCCT GACGTTGATG CTGCTGTCAT TCCTGGAAAT
 20 601 TTTGCTATAG CAGCAAACCT TTCTCCAAAG AAAGATAGTC TTTGTTTAGA
 651 GGATCTTTCC GTATCTAAGT ATACAAACCT TGTGTGCATT CGTTCTGAAG
 701 ACGTAGGTTT TCCTAAAATG ATAAAATTAC AGAAGCTGTT TCAATCTCCT
 751 TCTGTACAAC ATTTTTTTGA TACAAAATAT CATGGGAATA TTTTGACAAAT
 801 GACTCAAGAC AATGGTTAG

25 The PSORT algorithm predicts an inner membrane location (0.127).

The protein was expressed in *E. coli* and purified as a his-tag product, as shown in Figure 1A, and also as a GST-fusion. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 1B) and for FACS analysis (Figure 1C).

The cp6552 protein was also identified in the 2D-PAGE experiment (Cpn0278).

30 These experiments show that cp6552 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 2

The following *C. pneumoniae* protein (PID 4376736) was expressed <SEQ ID 3; cp6736>:

35 1 MKTSIRKFLI STTLAPCFAS TAFTVEVIMP SENFDGSSGK IFPYTTLSDP
 51 RGTLCIFSGD LYIANLDNAI SRTSSSCFSN RAGALQILGK GGVSFSLNIR
 101 SSADGAAISS VITQNPCLCP LSFSGFSQMI FDNCESLTSD TSASNVIPHA
 151 SAIYATPML FTNNDLSILFQ YNRSAGFGAA IRGTSITIEK TKKSLLFNGN
 201 GSISNGGALT GSAAINLINN SAPVIFSTNA TGIYGGAIYL TGGSMILTSGN
 251 LSGVLFVNNS SRSGGAIYAN GNVTFSSNSD LTFQNTASP QNSLPAPTTP
 40 301 PTPPAVTPLL GYGGAIFCTP PATPPPTGVS LTISGENSVT FLENIASEQG
 351 GALYGKKISI DSNKSTIFLG NTAGKGAIA IPESGELSLS ANQGDILFNK
 401 NLSITSGTPT RNSIHFGKDA KFATLGATQG YTLFYFDPTT SDDLAAASAA
 451 ATVVVNPKAS ADGAYSGTIV FSGETLTATE AATPANATST LNQKLELEGG
 501 TLALRNGATL NVHNFTQDEK SVVIMDAGTT LATTNGANNT DGATTLNKLK
 45 551 INLDLSDGTF AAVNVQSTN GALTISGTLG LVKNSQDCCD NHGMFNKDLQ
 601 QVPILELKAT SNTVTTTDFS LGTNGYQQSP YGYQGTWEFT IDTTTHTVTG
 651 NWKKTGYLPH PERLAPLIPN SLWANVIDLR AVSQASAADG EDVPGKQLSI
 701 TGITNFFHAN HTGDARSYRH MGGGYLINTY TRITPDAALS LGFGQLFTKS
 751 KDYLVGHHGS NVYFATVYSN ITKSLFGSSR FFSGGTSRVT YSRSEKVKKT
 50 801 SYTKLPKGRS SWSNNCWLGE LEGNLPITLS SRILNLKQII PFVKAEVAYA
 851 THGGIQENTP EGRIFGHGHL LNVAVPVGVR FGKNSHNRPD FYTIIIVAYAP
 901 DVYRHNPDCE TTLPIGATW TSGNNLTRS TLLVQASSHT SVNDVLEIFG
 951 HCGCDIRRTS RQYTLDIGSK LRF*

A predicted signal peptide is highlighted.

The cp6736 nucleotide sequence <SEQ ID 4> is:

```

1  ATGAAAACGT CTATTCGTAA GTTCTTAATT TCTACCACAC TGGCGCCATG
51  TTTTGCTTCA ACAGCGTTTA CTGTAGAAGT TATCATGCCT TCCGAGAACT
101  TTGATGGATC GAGTGGGAAG ATTTTTCCTT ACACAACACT TTCTGATCCT
151  AGAGGGACAC TCTGTATTTT TTCAGGGGAT CTCTACATTG CGAATCTTGA
201  TAATGCCATA TCCAGAACCT CTTCAGTTG CTTTAGCAAT AGGGCGGGAG
251  CACTACAAAT CTAGGAAAAA GGTGGGGTTT TCTCCTTCTT AAATATCCGT
301  TCTTCAGCTG ACGGAGCCGC GATTAGTAGT GTAATCACCC AAAATCTCTGA
351  ACTATGTCCC TTGAGTTTTC CAGGATTTAG TCAGATGATC TTCGATAACT
401  GTGAATCTTT GACTTCAGAT ACCTCAGCGA GTAATGTCAT ACCTCACGCA
451  TCGGCGATTT ACGCTACAAC GCCCATGCTC TTACAAAACA ATGACTCCAT
501  ACTATTCCAA TACAACCGTT CTGCAGGATT TGGAGCTGCC ATTCGAGGCA
551  CAAGCATCAC AATAGAAAAT ACGAAAAAGA GCCTTCTCTT TAATGGTAAT
601  GGATCCATCT CTAATGGAGG GGCCCTCACG GGATCTGCAG CGATCAACCT
15  651  CATCAACAAT AGCGCTCCTG TGATTTTCTC AACGAATGCT ACAGGGATCT
701  ATGGTGGGGC TATTTACCTT ACCGGAGGAT CTATGCTCAC CTCTGGGAAC
751  CTCTCAGGAG TCTTGTTCTG TAATAATAGC TCGCGCTCAG GAGGCGCTAT
801  CTATGCTAAC GGAAATGTCA CATTTTCTAA TAACAGCGAC CTGACTTTCC
851  AAAACAAATC AGCAATCTCA CAAAACCTCT TACCTGCACC TACACCTCCA
20  901  CCTACACCAC CAGCAGTCAC TCCTTTGTTA GGATATGGAG GCGCCATCTT
951  CTGTACTCCT CCAGCTACCC CCCCAACCA AGGTGTTAGC CTGACTATAT
1001  CTGGAGAAAA CAGCGTTACA TTCCTAGAAA ACATTGCCTC CGAACCAAGGA
1051  GGAGCCCTCT ATGGCAAAAA GATCTCTATA GATTCTAATA AATCTACAAT
1101  ATTTCTTGGA AATACAGCTG GAAAAGGAGG CGCTATTGCT ATTCCCGAAT
25  1151  CTGGGGAGCT CTCTCTATCC GCAAATCAAG GTGATATCCT CTTTAAACAAG
1201  AACCTCAGCA TCACTAGTGG GACACCTACT CGCAATAGTA TTCACTTCGG
1251  AAAAGATGCC AAGTTTGCCA CTCTAGGAGC TACGCAAGGC TATACCTAT
1301  ACTTCTATGA TCCGATTACA TCTGATGATT TATCTGCTGC ATCCGCAGCC
1351  GCTACTGTGG TCGTCAATCC CAAAGCCAGT GCAGATGGTG CGTATTCAGG
30  1401  GACTATTGTC TTTTCAGGAG AAACCTCAC TGCTACCGAA GCAGCAACCC
1451  CTGCAAAATG TACATCTACA TTAAACCAA AGCTAGAACT TGAAGCGGT
1501  ACTCTCGCTT TAAGAAACGG TGCTACCTTA AATGTTTATA ACTTCACGCA
1551  AGATGAAAAG TCCGTCGTCA TCATGGATGC AGGGACCACA TTAGCAACTA
35  1601  CAAATGGAGC TAATAATACT GACGGTGCTA TCACCTTAAA CAAGCTTGTA
1651  ATCAATCTGG ATTTCTTTGGA TGGCACTAAA GCGGCTGTGC TTAATGTGCA
1701  GAGTACCAAT GGAGCTCTCA CTATATCCGG AACTTTAGGA CTGTGAAAA
1751  ACTCTCAAGA TTGCTGTGAC AACCACGGGA TGTTTAATAA AGATTTACAG
1801  CAAGTTCCGA TTTTAGAACT CAAAGCGACT TCAAATACTG TAACCACTAC
40  1851  GGACTTCAGT CTCGGCACAA ACGGCTATCA GCAATCTCCC TATGGGTATC
1901  AAGGAAC'TTG GGAGTTTACC ATAGACACGA CAACCCATAC GGTACAGGA
1951  AA'TTGAAAAA AAACCGGTTA TCTTCCTCAT CCGGAGCGTC TTGCTCCCTT
2001  CATTCCTAAT AGCCTATGGG CAAACGTCAT AGATTTACGA GCTGTAAGTC
2051  AAGCGTCAGC AGCTGATGGC GAAGATGTCC CTGGAAGCA ACTGAGCATC
45  2101  ACAGGAATTA CAAATTTCTT CCATGCGAAT CATACCGGTG ATGACGCGAG
2151  CTACCGCCAT ATGGGTGGAG GCTACCTCAT CAATACCTAC ACACGCATCA
2201  CTCCAGATGC TCGTTAAGT CTAGGTTTTC GACAGCTGTT TACAAAATCT
2251  AAGGATTACC TCGTAGGTCA CGGTCATTCT AACGTTTATT TCGCTACAGT
2301  ATACTCTAAC ATCACCAGT CTCTGTTTGG ATCATCGAGA TTCTTCTCAG
50  2351  GAGGCACTTC TCGAGTTACC TATAGCCGTA GCAATGAGAA AGTAAAGACT
2401  TCATATACAA AATTGCCTAA AGGGCGCTGC TCTTGGAGTA ACAATTGCTG
2451  GTTAGGAGAA CTCGAAGGGA ACCTTCCCAT CACTCTCTCT TCTCGCATCT
2501  TAAACCTCAA GCAGATCATT CCCTTTGTAA AAGCTGAAGT TGCTTACGCG
2551  ACTCATGGGG GCATCCAAGA AAATACCCCG GAGGGGAGGA TTTTGGACA
55  2601  CGGTCACTTA CTCAACGTTG CAGTTCCCGT AGGCGTCCGC TTTGGTAAAA
2651  ATTCTCATAA TCGACCAGAT TTTTACATA TAATCGTAGC CTATGCTCCT
2701  GATGCTATC GTCACAATCC TGATTGCGAT ACGACATTAC CTATTAATGG
2751  AGCTACGTGG ACCTCTATAG GGAATAATCT AACCAGAAGT ACTTTGCTAG
2801  TACAAGCATC CAGCCATACT TCAGTAAATG ATGTTCTAGA GATCTTCGGG
2851  CACTGTGGAT GTGATATTTC CAGAACCCTC CGTCAATATA CTCTAGATAT
60  2901  AGGAAGCAAA TTACGATTTT AA

```

The PSORT algorithm predicts an outer membrane location (0.917).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 2A, and also as a GST-fusion. Both proteins were used to immunise mice, whose sera were used in a Western blot (Figure 2B) and for FACS analysis (Figure 2C).

The cp6736 protein was also identified in the 2D-PAGE experiment (Cpn0453) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6736 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 3

The following *C.pneumoniae* protein (PID 4376751) was expressed <SEQ ID 5; cp6751>:

```

10      1  MRFFCFGMLL  PFTFVLANEG  LQLPLETYIT  LSPEYQAAPO  VGFTHNQNDQ
      51  LAIVGNHND  ILDYKYRSN  GGALTCKNLL  ISENIGNVFF  EKNVCPNSGG
     101  AIYAAQNCTI  SKNQNYAFTT  NLVSDNPTAT  AGSLLGGALF  AINCISITNNL
     151  GQGTFDNLA  LNKGGALYTE  TNLSEKDNKG  PIIKQNRAL  NSDSLGGGIY
     201  SGNLSNIEGN  SGAIQITSNS  SGSGGGIFST  QLTITSSNKK  LIEISENSAF
     15      251  ANNYGSNFNP  GGGGLTTTFC  TILNNREGVL  FNNNQSQSNG  GAIHAKSIII
     301  KENGVPYFLN  NTATRGGALL  NLSAGSGNGS  FILSADNGDI  IFNNNTASKH
     351  ALNPPYRNAI  HSTPNMNLQI  GARPGYRVLF  YDPIEHELPS  SFPILNFET
     401  GHTGTVLFSG  EHVHQNFTDE  MNFFSYLRNT  SELRQGVLA  V  EDGAGLACYK
     451  FFQRGGTL  LL  GQGA  VITTAG  TIPTPSSTPT  TVGSTITL  NH  I  AIDLPSILS
     20      501  FQAQAPKIWI  YPTKTGSTYT  EDSNPTITIS  GTLTLRNSN  N  EDPYDSL  DLS
     551  HSLEKVPLLY  IVDVAAQKIN  SSQDLSTLN  SGHEGYQGI  WSTYVWETTT
     601  ITNPTSL  LGA  NTKHKL  LLYAN  WSPLGYR  PHP  ERRGEFIT  NA  LWQSAYTA  LA
     651  GLHSLSSW  DE  EKGHAAS  LQG  IGLLVH  QKDK  NGFKGFR  SHM  TGYSATTE  AT
     701  SSQSPNFS  LG  FAQFFS  KAKE  HESQNST  SSH  HYFSGMC  IEN  TLFKEWIR  LS
     25      751  VSLAYMFT  SE  HTHTMY  QGLL  EGNSQGS  FHN  HTLAGAL  SCV  FLPQPHGE  SL
     801  QIYPFIT  ALA  IRGNLA  AFQE  SGDHARE  FSL  HRPLTDV  SLP  VGI  RASWKNH
     851  HRVPLV  WLTE  ISYRST  LYRQ  DPELH  SKLLI  SQGTWT  TQAT  PVTYNAL  GIK
     901  VKNTMQV  FPK  VTLSLD  YSAD  ISSSTL  SHYL  NVASRM  RF*

```

A predicted signal peptide is highlighted.

30 The cp6751 nucleotide sequence <SEQ ID 6> is:

```

      1  ATGCGCTTTT  TTTGCTTCGG  AATGTTGCTT  CCTTTTACTT  TTGTATTGGC
     51  TAATGAAGGT  CTCCAACCTC  CTTTGGAGAC  CTATATTACA  TTAAGTCCTG
    101  AATATCAAGC  AGCCCCTCAA  GTAGGGTTTA  CTCATAACCA  AAATCAAGAT
    151  CTCGCAATTG  TCGGGAATCA  CAATGATTTT  ATCTTGGACT  ATAAGTACTA
    201  TCGGTGCAAT  GGAGGTGCTC  TTACCTGTAA  GAATCTTCTG  ATCTCTGAAA
    251  ATATAGGGAA  TGTCTTCTTT  GAGAAGAATG  TCTGTCCCAA  TTCTGGCGGG
    301  GCAATTATG  CTGCTCAAAA  TTGCACGATC  TCCAAGAATC  AGAATATATG
    351  ATTTACTACA  AACTTGGTCT  CTGACAATCC  TACAGCCACT  GCGGGATCAC
    401  TATTGGGTGG  AGCTCTCTTT  GCCATAAATT  GCTCTATTAC  TAATAACCTA
    451  GGACAGGGAA  CTTTCGTTGA  CAATCTCGCT  TTAAATAAGG  GGGGTGCCCT
    501  CTATACTGAG  ACGAACTTAT  CTATTAAAGA  CAATAAAGGC  CCGATCATAA
    551  TCAAGCAGAA  TCGGGCACTA  AATTCGGACA  GTTTAGGAGG  AGGGATTTAT
    601  AGTGGGAAC  TCTTAAATAT  AGAGGGAAT  TCTGGAGCTA  TACAGATCAC
    651  AAGCAACTCT  TCAGGATCTG  GGGGAGGCAT  ATTTTCTACC  CAAACACTCA
    701  CGATCTCCTC  GAATAAAAAA  CTCATAGAAA  TCAGTGAAAA  TTCCGCGTTC
    751  GCAATAACT  ATGGATCGAA  CTTCAATCCA  GGAGGAGGAG  GTCCTTACTAC
    801  CACCTTTTGC  ACGATATTGA  ACAACCGAGA  AGGGGTACTC  TTTAACAATA
    851  ACCAAAGCCA  GAGCAACGGT  GGAGCCATTC  ATGCGAAATC  TATCATTATC
    901  AAAGAAAATG  GTCCTGTATA  CTTTTTAAAT  AACACTGCAA  CTCGGGGAGG
    951  GGCTCTCCTC  AACTTATCAG  CAGGTTCTGG  AAACGGAAGC  TTCATCTTAT
   1001  CTGCAGATAA  TGGAGATATT  ATCTTTAACA  ATAATACGGC  CTCCAAGCAT
   1051  GCCCTCAATC  CTCCATACAG  AAACGCCATT  CACTCGACTC  CTAATATGAA
   1101  TCTGCAAATA  GGAGCCCGTC  CCGGCTATCG  AGTGCTGTTC  TATGATCCCA
   1151  TAGAACATGA  GCTCCCTTCC  TCCTTCCCCA  TACTCTTTAA  TTTTCGAAAC
   1201  GGTCAATACG  GTACAGTTTT  ATTTTCAGGG  GAACATGTAC  ACCAGAACTT

```

	1251	TACCGATGAA	ATGAATTTCT	TTTCTTATTT	AAGGAACACT	TCGGAACACT
	1301	GTCAGGAGT	CCTTGCTGTT	GAAGATGGTG	CGGGGCTGGC	CTGCTATAAG
	1351	TTCTTCCAAC	GAGGAGGCAC	TCTACTTCTA	GGTCAAGGTG	CGGTGATCAC
5	1401	GACAGCAGGA	ACGATTCCCA	CACCATCCTC	AACACCAACG	ACAGTAGGAA
	1451	GTACTATAAC	TTTAAATCAC	ATTGCCATTG	ACCTTCCTTC	TATTCTTTCT
	1501	TTTCAAGCTC	AGGCTCCAAA	AATTTGGATT	TACCCACAA	AAACAGGATC
	1551	TACCTATACT	GAAGATTCCA	ACCCGACAAT	CACAATCTCA	GGAACCTCTCA
	1601	CCTTACGCAA	CAGCAACAAC	GAAGATCCCT	ACGATAGTCT	GGATCTCTCG
10	1651	CACCTCTCTG	AGAAAGTTCC	CCTTCTTTAT	ATTGTCGATG	TCGCTGCACA
	1701	AAAAATTAAC	TCTTCGCAAC	TGGATCTATC	CACATTAAT	TCTGGCGAAC
	1751	ACTATGGGTA	TCAAGGCATC	TGGTCGACCT	ATTGGGTAGA	AACTACAACA
	1801	ATCACGAACC	CTACATCTCT	ACTAGGCGCG	AATACAAAAC	ACAAGCTGCT
	1851	CTATGCAAAC	TGGTCTCCTC	TAGGCTACCG	TCCTCATCCC	GAACGTCGAG
	1901	GAGAATTCAT	TACGAATGCC	TTGTGGCAAT	CGGCATATAC	GGCTCTTGCA
15	1951	GGACTCCACT	CCCTCTCCTC	CTGGGATGAA	GAGAAGGGTC	ATGCAGCTTC
	2001	CCTACAAGGC	ATTGGTCTTC	TGGTTCATCA	AAAAGACAAA	AACGGTTTTA
	2051	AGGGATTTTC	TAGTCATATG	ACAGGTTATA	GTGCTACCAC	CGAAGCAACC
	2101	TCTTCTCAA	GTCCGAATTT	CTCTTTAGGA	TTTGCTCAGT	TCTTCTCCAA
20	2151	AGCTAAAGAA	CATGAATCTC	AAAATAGCAC	GTCCCTTCAC	CACATTTTCT
	2201	CTGGAATGTG	CATAGAAAAT	ACTCTCTTCA	AAGAGTGGAT	ACGTCTATCT
	2251	GTGTCTCTTG	CTTATATGTT	TACCTCGGAA	CATACCCATA	CAATGTATCA
	2301	GGGTCTCCTG	GAAGGGAAC	CTCAGGGATC	TTTCCACAAC	CATACCTTAG
	2351	CAGGGCTCT	CTCCTGTGTT	TTCTTACCTC	AACCTCACGG	CGAGTCCCTG
25	2401	CAGATCTATC	CCTTTATTAC	TGCCCTAGCC	ATCCGAGGAA	ATCTTGCTGC
	2451	GTTTCAAGAA	TCTGGAGACC	ATGCTCGGGA	ATTTTCCCTA	CACCGCCCCC
	2501	TAACGGACGT	CTCCCTCCCT	GTAGGAATCC	GCGCTTCTTG	GAAGAACCAC
	2551	CACCGAGTTC	CCCTAGTCTG	GCTCACAGAA	ATTCCTATC	GCTCTACTCT
	2601	CTATAGGCAA	GATCCTGAAC	TCCACTCGAA	ATTACTGATT	AGCCAAGGTA
30	2651	CGTGGACGAC	GCAGGCCACT	CCTGTGACCT	ACAAATGCTT	AGGGATCAAA
	2701	GTGAAAAATA	CCATGCAGGT	GTTTCCTAAA	GTCACTCTCT	CCTTAGATTA
	2751	CTCTGCGGAT	ATTTCTTCCT	CCACGCTGAG	TCACTACTTA	AACGTGGCGA
	2801	GTAGAATGAG	ATTTTAA			

The PSORT algorithm predicts an outer membrane location (0.923).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 3A, and also in his-tagged form. The GST-fusion recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 3B) and for FACS analysis (Figure 3C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6751 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 4

The following *C.pneumoniae* protein (PID 4376752) was expressed <SEQ ID 7; cp6752>:

	1	MFGMTPAVYS	LQTDLSLEKFA	LERDEEFRTS	FPLLDLSLSTL	TGFSPITTFV
	51	GNRHNSSQDI	VLSNYKSIDN	ILLLWTSAGG	AVSCNNFLLS	NVEDHAFFSK
45	101	NLAIGTGGA	ACQGAQTITK	NRGPLIFFSN	RGLNNASTGG	ETRGGAIAACN
	151	GDFTISQNG	TFYFVNNSVN	NWGGALSTNG	HCRIQSNRAP	LLFFNNTPAS
	201	GGGALRSNT	TIISDNTRPIY	FKNNCGNNGG	AIQTSVTVAI	KNNSGSVIFN
	251	NNFALSGSIN	SGNGSGGAIY	TTNLSIDNPN	GTILFNNNYC	IRDGGAICTQ
	301	PLTIKNSGHV	YFTNNQGNWG	GALMLLQDST	CLLFABEQNI	AFQNNNEVFLT
50	351	TFGRYNAIHC	TPNSNLQLGA	NKGYTTAFD	PIEHQHPTTN	PLIFNPANAN
	401	QGTILFSSAY	IPEASDYENN	FISSSKNTSE	LRNGVLSIED	RAGWQFYKFT
	451	QKGGILKLGH	AASIATTANS	ETPSTSVGSQ	VIINNLAINL	PSILAKGKAP
	501	TLWIRPLQSS	APFTEDNNPT	ITLSGPLTLL	NEENRDPYDS	IDLSEPLQNI
	551	HLLSLSDVTA	RHINTDNFHP	ESLNATEHYG	YQGIWSPYVW	ETITTTNNAS
55	601	IETANTLYRA	LYANWTPLGY	KVNPEYQGD	ATTPLWQSFH	TMFSLLRYSN
	651	RTGDSIERP	FLEIQGIADG	LFVHQNSIPG	APGFRIQSTG	YSLQASSETS

5 701 LHQKISLGFA QFFTRTKEIG SSMNVSAHNT VSSLYVELPW FQEFATSTV
 751 LAYGYGDHHL HSLHPSHQEQ AEGTCYSHTL AAAIGCSFPW QOKSYLHLSP
 801 FVQAIARSH QTAFFEEIGDN PRKFVSQKPF YNLTLPLGIQ GKWQSKFHPV
 851 TEWTLLELSYQ PVLYQQNPQI GVTLLASGGS WDILGHNYVR NALGYKVHNQ
 901 TALFRSLDLF LDYQGSVSSS TSTHHLQAGS TLKF*

The cp6752 nucleotide sequence <SEQ ID 8> is:

1 ATGTTTCGGGA TGACTCCTGC AGTGATAGT TTACAAACGG ACTCCCTTGA
 51 AAAGTTTGCT TTAGAGAGGG ATGAAGAGTT TCGTACGAGC TTTCCTCTCT
 10 101 TAGACTCTCT CTCCACTCTT ACAGGATTTT CTCCAATAAC TACGTTTGTT
 151 GGAAATAGAC ATAATTCCTC TCAAGACATT GTACTTCTA ACTACAAGTC
 201 TATTGATAAC ATCCTTCTTC TTTGGACATC GGCTGGGGGA GCTGTGTCTT
 251 GTAATAATTT CTTATTATCA AATGTTGAAG ACCATGCCTT CTTCAGTAAA
 301 AATCTCGCGA TTGGGACTGG AGCGCGGATT GCTTGCCAGG GAGCCTGCAC
 351 AATCACGAAG AATAGAGGAC CCCTTATTTT TTTCAGCAAT CGAGGTCTTA
 401 ACAATCGCAG TACAGAGGGA GAACTCGTG GGGGTGCGAT TGCTTCAACA
 451 GGAGACTTCA CGATTCTCA AATCAAGGG ACTTCTACT TTGTCAACAA
 501 TTCCGTCAAC AACTGGGGAG GAGCCTCTC CACCAATGGA CACTGCCGCA
 551 TCCAAAGCAA CAGGGCACCT CTACTCTTTT TTAACAATAC AGCCCTTAGT
 601 GGAGGGGGTG CGCTTCGTAG TGAAAATACA ACGATCTCTG ATAACACGCG
 651 TCCTATTTAT TTTAAGAACA ACTGTGGGAA CAATGGCGGG GCCATTCAAA
 701 CAAGCGTTAC TGTTCGATA AAAAATAACT CCGGTTCGGT GATTCTCAAT
 751 AACACACAG CGTTATCTGG TTCGATAAAT TCAGGAAATG GTTCAGGAGG
 801 GGCGATTAT ACAACAACC TATCCATAGA CGATAACCCT GGAATATTTC
 851 TTTTCAATAA TAACCTACTG ATTCCGATG GCGGAGCTAT CTGTACACAA
 901 TTTTGCATA TCAAAAATAG TGGCCACGTA TATTTACCA ACAATCAAGG
 951 AAACGTGGGA GGTGCTCTTA TGCTCCTACA GGACAGCACC TGCTACTCT
 1001 TCGCGGAACA AGGAAATATC GCATTCAAA ATAATGAGGT TTTCCTCACC
 1051 ACATTTCGTA GATACAACGC CATACTTGT ACACCAATA GCAACTTACA
 1101 ACTTGAGCTT AATAAGGGGT ATACGACTGC TTTTTCGAT CCTATAGAAC
 1151 ACCAATATCC AACTACAAAT CCTCTAATCT TTAATCCCAA TGCGAACCAT
 1201 CAGGGAACGA TCTTATTTTC TTCAGCCTAT ATCCAGAAG CTTCTGACTA
 1251 CGAAAATAAT TTCAATAGCA GCTCGAAAAA TACCTCTGAA CTTCCGAATG
 1301 GTGTCTCTCT TATCGAGGAT CGTGCGGGAT GGCAATCTA TAAGTTCAC
 1351 CAAAAAGGAG GTATCCTTAA ATTAGGGCAT GCGGCGAGTA TTGCAACAAC
 1401 TGCCAACTCT GAGACTCCAT CAACTAGTGT AGGCTCCAG GTCATCATTA
 1451 ATAACCTTGC GATTAACTC CCCTCGATCT TAGCAAAAGG AAAAGCTCCT
 1501 ACCTTGTTGA TCCGTCTCTT ACAATCTAGT GCTCCTTTCA CAGAGGACAA
 1551 TAACCTTACA ATTACTTTAT CAGGTCTCTT GACACTCTTA AATGAGGAAA
 1601 ACCGCGATCC CTACGACAGT ATAGATCTCT CTGAGCCTTT ACAAAACATT
 1651 CATCTTCTTT CTTTATCGGA TGTAAACAGCA CGTCATATCA ATACCGATAA
 1701 CTTTCATCCT GAAAGCTTAA ATGCGACTGA GCATTACGGT TATCAAGGCA
 1751 TCTGCTCTCC TTATTGGGTA GAGACGATAA CAACAACAAA TAACGCTTCT
 1801 ATAGAGACGG CAAACACCCCT CTACAGAGCT CTGTATGCCA ATTGGACTCC
 1851 CTTAGGATAT AAGGTCAATC CTGAATACCA AGGAGATCTT GCTACGACTC
 1901 CCTATGGCA ATCCTTTTCA CTATGTTCT CTCTATTAAG AAGTTATAAT
 1951 CGAACTGGTG ATTCTGATAT CGAGAGGCCT TTCTTAGAAA TTCAGGGAT
 2001 TGCCGACGGC CTCTTTGTTT ATCAAAATAG CATCCCCGGG GCTCCAGGAT
 2051 TCCGTATCCA ATCTACAGG TATTCCTTAC AAGCATCCCT CGAACTTCT
 2101 TTACATCAGA AAATCTCCTT AGGTTTTCGA CAGTTCTTCA CCCGCACTAA
 2151 AGAAATCGGA TCAAGCAACA ACGTCTCGGC TCACAATACA GTCCTCTCAC
 2201 TTTATGTTGA GCTTCCGTGG TTCCAAGAGG CTTTGCAC ATCCACAGTG
 2251 TTAGCGTATG GCTATGGGGA CCATCACCTC CACAGCCTAC ATCCCTCACA
 2301 TCAAGAACAG GCAGAAGGGA CGTGTATAG CCATACATTA GCAGCAGCTA
 2351 TCGGCTGTTT TTTCCCTTGG CAACAGAAAT CCTATCTTCA CCTCAGCCCG
 2401 TTCGTTACAG CAATTGCAAT ACGTCTCTAC CAAACAGCGT TCGAAGAGAT
 2451 TGGTGACAA CCCCAGAAAT TTGTCTCTCA AAAGCCTTTC TATAATCTGA
 2501 CCTTACCTCT AGGAATCCAA GGAAATGGC AGTCAAAATT CCACGTACCT
 2551 ACAGATGGA CTCTAGAACT TTCTTACCAA CCGTACTCT ATCAACAAAA
 2601 TCCCCAAATC GGTGTCACGC TACTTGCGAG CGGAGGTTCC TGGGATATCC
 2651 TAGGCCATAA CTATGTTTCG AATGCTTTAG GGTACAAAGT CCACAATCAA
 2701 ACTGCGCTCT TCCGTTCTCT CGATCTATTC TTGGATTACC AAGGATCGGT
 2751 CTCCTCCTCG ACATCTACGC ACCATCTCCA AGCAGGAAGT ACCTTAAAAA
 2801 TCTAA

The PSORT algorithm predicts a cytoplasmic location (0.138).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 4A, and also as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (4B) and the his-tagged protein was used for FACS analysis (4C).

The cp6752 protein was also identified in the 2D-PAGE experiment (Cpn0467).

- 5 These experiments show that cp6752 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 5

The following *C.pneumoniae* protein (PID 4376850) was expressed <SEQ ID 9; cp6850>:

10 1 MKRAVLIAAM FCGVVSLSSC CRIVDCCFED PCAPSSCNPC EVIRKKERSC
51 GGNACGSYVP SCSNPGSGTE CNSQSPQVKG CTSPDGRCKQ *

A predicted signal peptide is highlighted.

The cp6850 nucleotide sequence <SEQ ID 10> is:

1 ATGAAGAAAG CTGTTTTAAT TGCTGCAATG TTTTGTGGAG TAGTTAGCTT
15 51 AAGTAGCTGC TGCCGCATTG TAGATTGTTG TTTTGAGGAT CCTTGCGCAC
101 CCTCTTCTTG CAATCCTTGT GAAGTAATAA GAAAAAAGA AAGATCCTGC
151 GCGGTAATG CTTGTGGGTC CTACGTTTCT TCTTGTTCCT ATCCATGTGG
201 TTCAACAGAG TGTAACCTCT AAAGCCACCA AGTTAAAGGT TGTACATCAC
251 CTGATGGCAG ATGCAAACAG TAA

The PSORT algorithm predicts an inner membrane location (0.329).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 5A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 5B) and for FACS analysis (Figure 5B). A his-tagged protein was also expressed.

These experiments show that cp6850 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

25 Example 6

The following *C.pneumoniae* protein (PID 4376900) was expressed <SEQ ID 11; cp6900>:

1 MKIKFSWKVN FLICLLAVGL IFFGCSRVRK EVLVGRDATW FPKQFGIYTS
51 DTNAFLNDLV SEINYKENLN INIVNQDWVH LFENLDDKKT QGAFTSVLPT
30 101 LEMLEHYQFS DPILLTGPVL VVAQDSPYQS IEDLKGRLIG VYKFDSSVLV
151 AQNIPDAVIS LYQHVPIALE ALTSNCYDAL LAPVIEVTAL IETAYKGRLE
201 IISKPLNADG LRLAILKGFTN GDLLEGFNAG LVKTRRSRGY DAIKQRYRLP

The cp6900 nucleotide sequence <SEQ ID 12> is:

1 GTGAAGATAA AATTTTCTTG GAAGGTAAAT TTTTAAATAT GTTTACTGGC
35 51 TGTGGGACTG ATCTTTTTCG GGTGCTCTCG AGTAAAAAGA GAAGTTCTCG
101 TAGGTCGTGA TGCCACCTGG TTTCCAAAAC AATTCGGCAT TTATACATCC
151 GATACCAACG CATTTTAAAC CGATCTTGTG TCTGAGATTA ACTATAAGA
201 GAATCTAAAT ATTAATATATG TAAATCAAGA TTGGGTGCAT CTCTTTGAGA
251 ATTTAGATGA TAAAAAGACC CAAGGAGCAT TTACATCTGT ATTGCCCTACT
301 CTGAGATGC TCGAACACTA TCAATTTTCT GATCCCATTT TACTCACAGG
40 351 TCCTGTCTT GTCGTCGCTC AAGACTCTCC TTACCAATCT ATAGAGGATC
401 TTAAAGGTCG TCTTATTGGA GTGTATAAGT TTGACTCTTC AGTTCTTGTA
451 GCTCAAAATA TCCCTGACGC TGTGATTAGC CTCTACCAAC ATGTTCCAAT
501 AGCATTTGGA GCCTTAACAT CGAATTGTTA CGACGCTCTT CTAGCTCCTG
551 TAATTGAAGT GACCGCGCTA ATAGAAACAG CATATAAAGG AAGACTGAAA
45 601 ATTATTTCAT AACCTTTAAA CGCAGATGGT TTGCGGCTTG CAATACTGAA

651 AGGGACAAAC GGAGATTGTC TTGAAGGGTT TAACGCAGGA CTGTGAAAAA
 701 CACGACGCTC AGGAAAATAC GATGCTATAA AACAGCGGTA TCGTCTTCCC
 751 TAA

The PSORT algorithm predicts an inner membrane location (0.452).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 6A. The recombinant protein was used to immunise mice, whose sera were used for FACS analysis (Figure 6B). A his-tagged protein was also expressed.

The cp6900 protein was also identified in the 2D-PAGE experiment (Cpn0604).

- 10 These experiments show that cp6900 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 7

The following *C.pneumoniae* protein (PID 4377033) was expressed <SEQ ID 13; cp7033>:

15
 20
 25

1 MVNPIGPGPI DETERTPPAD LSAQGLEASA ANKSAEAQRI AGAEAKPKES
 51 KTDSVERWSI LRSVAVNALMS LADKLGILASS NSSSSTSRSA DVDSTTATAP
 101 TPPPPTFDDY KTQAQTAYDT IFTSTSLADI QAALVSLQDA VTNIKDTAAT
 151 DEETAIAAEW ETKNADAVKV GAQITELAKY ASDNQAILDS LGKLTSTFDLL
 201 QAALLQSVAN NNKAAELLKE MQDNPVVPFK TPAIAQSLVD QTDATATQIE
 251 KDGNAIIRDY FAGQNASGAV ENAKSNNSIS NIDSAKAAIA TAKTQIAEAQ
 301 KKFDPSPILQ EAEQMVIAE KDLKNIKPAD GSDVPNPGTT VGGSKQQGSS
 351 IGSIRVSMLL DDAENETASI LMSGFRQMIH MFNTENPDSQ AAQQLAAQA
 401 RAAKAAGDDS AAAALADAQK ALEAALGKAG QQQGILNALG QIASAAVVSA
 451 GVPPAAASSI GSSVKQLYKT SKSTGSDYKT QISAGYDAYK SINDAYGRAR
 501 NDATRDVINN VSTPALTRSV PRARTEARGP EKTDQALARV ISGNSRTLGD
 551 VYSQVSALQS VMQIIQSNPQ ANNEEIRQKL TSAVTKPPQF GYPYVQLSND
 601 STQKFIAKLE SLFAEGSRTA AEIKALSFET NSLFIQQVLV NIGSLYSGYL
 651 Q*

The cp7033 nucleotide sequence <SEQ ID 14> is:

30
 35
 40
 45
 50
 55

1 ATGGTTAATC CTATTGGTCC AGGTCCTATA GACGAAACAG AACGCACACC
 51 TCCCGCAGAT CTTTCTGCTC AAGGATTGGA GGCGAGTGCA GCAAATAAGA
 101 GTGCGGAAGC TCAAAGAATA GCAGGTGCGG AAGCTAAGCC TAAAGAATCT
 151 AAGACCGATT CTGTAGAGCG ATGGAGCATC TTGCGTCTTG CAGTGAATGC
 201 TCTCATGAGT CTGGCAGATA AGCTGGGTAT TGCTTCTAGT AACAGCTCGT
 251 CTTCTACTAG CAGATCTGCA GACGTGGACT CAACGCAGC GACCGCACCT
 301 ACGCCTCCTC CACCCACGTT TGATGATTAT AAGACTCAAG CGCAACAGC
 351 TTACGATACT ATCTTTACCT CAACATCACT AGCTGACATA CAGGCTGCTT
 401 TGGTGAGCCT CCAGGATGCT GTCACATAA TAAAGGATAC AGCGGCTACT
 451 GATGAGGAAA CCGCAATCGC TGCGGAGTGG GAAACTAAGA ATGCCGATGC
 501 AGTTAAAGTT GGC CGC GCAAA TTACAGAATT AGCGAAATAT GCTTCGGATA
 551 ACCAAGCGAT TCTTGACTCT TTAGGTAAAC TGA CTTCCTT CGACCTCTTA
 601 CAGGCTGCTC TTCTCCAATC TG TAGCAAAC AATAACAAAG CAGCTGAGCT
 651 TCTTAAAGAG ATGCAAGATA ACCCAGTAGT CCCAGGGAAG ACGCTGCAA
 701 TTGCTCAATC TTTAGTTGAT CAGACAGATG CTACAGCGAC ACAGATAGAG
 751 AAAGATGGAA ATGCGATTAG GGATGCATAT TTTGCAGGAC AGAACGCTAG
 801 TGGAGCTGTA GAAAATGCTA AATCTAATAA CAGTATAAGC AACATAGATT
 851 CAGCTAAAGC AGCAATCGCT ACTGCTAAGA CACAAATAGC TGAAGCTCAG
 901 AAAAAGTTCC CCGACTCTCC AATTCCTTCA GAAGCGGAAC AAATGGTAAT
 951 ACAGGCTGAG AAAGATCTTA AAAATATCAA ACCTGCAGAT GGTTCGTATG
 1001 TTCCAAATCC AGGAACTACA GTTGAGAGCT CCAAGCAACA AGGAAGTAGT
 1051 ATTGGTAGTA TTCGTGTTTC CATGCTGTTA GATGATGCTG AAAATGAGAC
 1101 CGTTCCATT TTGATGTCTG GGTTCGTCA GATGATTCAC ATGTTCAATA
 1151 CGGAAAATCC TGATTTCAA GCTGCCCAAC AGGAGCTCGC AGCACAAGCT
 1201 AGAGCAGCGA AAGCCGCTGG AGATGACAGT GCTGCTGCAG CGCTGGCAGA
 1251 TGCTCAGAAA GCTTTAGAAG CGGCTCTAGG TAAAGCTGGG CAACACAGG
 1301 GCATACTCAA TGCTTTAGGA CAGATCGCTT CTGCTGCTGT TGTGAGCGCA
 1351 GGAGTTCCTC CCGCTGCAGC AAGTTCCTATA GGGTCATCTG TAAACAGCT
 1401 TTACAAGACC TCAAAATCTA CAGGTTCCTGA TTATAAAACA CAGATATCAG

1451 CAGGTTATGA TGCTTACAAA TCCATCAATG ATGCCTATGG TAGGGCACGA
 1501 AATGATGCGA CTCGTGATGT GATAAACAAAT GTAAGTACCC CCGCTCTCAC
 1551 ACGATCCGTT CCTAGAGCAC GAACAGAAGC TCGAGGACCA GAAAAACAG
 1601 ATCAAGCCCT CGCTAGGGTG ATTTCTGGCA ATAGCAGAAC TCTTGGAGAT
 1651 GTCTATAGTC AAGTTTCGGC ACTACAATCT GTAATGCAGA TCATCCAGTC
 1701 GAATCCTCAA GCGAATAATG AGGAGATCAG ACAAAGCTT ACATCGGCAG
 1751 TGACAAAGCC TCCACAGTTT GGCTATCCTT ATGTGCAACT TTCTAATGAC
 1801 TCTACACAGA AGTTCATAGC TAAATTAGAA AGTTTGTTCG CTGAAGGATC
 1851 TAGGACAGCA GCTGAAATAA AAGCACTTTC CT'TTGAAACG AACTCCTTGT
 1901 TTATTCAGCA GGTGCTGGTC AATATCGGCT CTCTATATTC TGGTTATCTC
 1951 CAATAA

The PSORT algorithm predicts a cytoplasmic location (0.272).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 7A. A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used for FACS (Figure 7B) and Western blot (7C) analyses.

The cp7033 protein was also identified in the 2D-PAGE experiment (Cpn0728) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7033 a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 8

The following *C.pneumoniae* protein (PID 6172321) was expressed <SEQ ID 15; cp0017>:

1 MGIKGTGIIV WVDDATAKTK NATLTWPKTG YKPNPERQGP LVPNSLWGSF
 51 VDVRSIQSLM DRSTSSLSSS TNLWVSGIAD FLHEDQKGNQ RSYRHSSAGY
 101 ALGGGFFTAS ENFFNFAFQCQ LFGYDKDHLV AKNHTHVYAG AMSYRHLGES
 151 KTLAKILSGN SDSLPFVFNA RFAYGHTDNN MTTKYTGYSF VKGSWGNDAF
 201 GIECGGAIPV VASGRRSWVD THTPFLNLEM IYAHQNDFFE NGTEGRSFQS
 251 EDLFNLAVPV GIKFEKFSK STYDLAIAYV PDVIRNDPGC TTTLMVSGDS
 301 WSTCGTSLSR QALLVRAGNH HAFASNFEVF SQFEVELRGS SRSYIDLGG
 351 RFGF*

The cp0017 nucleotide sequence <SEQ ID 16> is:

1 ATGGGTATCA AGGGAACCTGG AATAATTGTT TGGGTCGACG ATGCAACTGC
 51 AAAAAACAAA AATGCTACCT TAAC'TTGGAC TAAAACAGGA TACAAGCCGA
 101 ATCCAGAACG TCAGGACCTT TTGGTTCCTA ATAGCCTGTG GGGTTC'TTTT
 151 GTCGATGTCC GCTCCATTCA GAGCCTCATG GACCGGAGCA CAAGTTCGTT
 201 ATCTTCGTCA ACAAATTTGT GGGTATCAGG AATCGCGGAC TTTT'TGCATG
 251 AAGATCAGAA AGGAAACCAA CGTAGTTATC GTCATTCTAG CGCGGGTTAT
 301 GCATTAGGAG GAGGATTCCT CACGGCTTCT GAAAA'TTCT TTAATTTTGC
 351 TTTT'TGTCAG CTTT'TTGGCT ACGACAAGGA CCATCTGTG GCTAAGAACC
 401 ATACCCATGT ATATGCAGGG GCAATGAGTT ACCGACACCT CGGAGAGTCT
 451 AAGACCCTCG CTAAGATTTT GTCAGGAAAT TCTGACTCCC TACCTTTTGT
 501 CTTCAATGCT CGGTTTGCTT ATGGCCATAC CGACAATAAC ATGACCACAA
 551 AGTACACTGG CTATTCTCCT GTTAAGGGAA GCTGGGGAAA TGATGCCTTC
 601 GGTATAGAAT GTGGAGGAGC TATCCCGGTA GTTGCTTCAG GACGTCGGTC
 651 TTGGGTGGAT ACCCACACGC CATTTCTAAA CCTAGAGATG ATCTATGCAC
 701 ATCAGAATGA CTTT'TAGGAA AACCGCACAG AAGGCCGTTT TTTCCAAAGT
 751 GAAGACCTCT TCAATCTAGC GGTTCCTGTA GGGATAAAAT TTGAGAAATT
 801 CTCCGATAAG TCTACGTATG ATCTCTCCAT AGCTTACGTT CCCGATGTGA
 851 TTCGTAATGA TCCAGGCTGC ACGACAATC TTATGGTTTC TGGGGATTCT
 901 TGGTCGACAT GTGGTACAAG CTTGTCTAGA CAAGCTCTTC TTGTACGTGC
 951 TGGAATCAT CATGCCTTTG CTTCAAACCT TGAAGTTTTC AGTCAGTTTG
 1001 AAGTCGAGTT GCGAGTTTCT TCTCGTAGCT ATGCTATCGA TCTTGGAGGA
 1051 AGATTCCGAT TTAA

This sequence is frame-shifted with respect to cp0016.

The PSORT algorithm predicts a cytoplasmic location (0.075).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 8A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 8B) and for FACS analysis (Figure 8C). A his-tagged protein was also expressed.

- 5 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0017 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 9

- 10 The following *C.pneumoniae* protein (PID 6172315) was expressed <SEQ ID 17; cp0014>:

```

1  MKSSFPKFVF STFAIFPLSM IATETVLDSS ASFDGNKNGN FSVRESQEDA
51  GTTYLFKGNV TLENIPGTGT AITKSCFNNT KGDLTFTGNG NSLLFQTVDA
101 GTVAGAAVNS SVVDKSTTFI GFSSLSFIAS PGSSITTGKG AVSCSTGSLM
151 LTKMSVCSSA KTFQRIMAVL SPQKLFH*
```

- 15 The cp0014 nucleotide sequence <SEQ ID 18> is:

```

1  ATGAAGTCTT CTTTCCCCAA GTTTGTATTT TCTACATTG CTATTTTCCC
51  TTGTCTATG ATTGCTACCG AGACAGTTTT GGATTCAAGT GCGAGTTTCG
101 ATGGGAATAA AAATGGTAAT TTTTCAGTTC GTGAGAGTCA GGAAGATGCT
151 GGAAGTACCT ACCTATTTAA GGGAAATGTC ACTCTAGAAA ATATTCCTGG
20 201 AACAGGCACA GCAATCACAA AAAGCTGTTT TAACAACACT AAGGGCGATT
251 TGACTTTCAC AGGTAACGGG AACTCTCTAT TGTTCCAAAC GGTGGATGCA
301 GGGACTGTAG CAGGGGCTGC TGTTAACAGC AGCGTGGTAG ATAAATCTAC
351 CACGTTTATA GGGTTTCTTT CGCTATCTTT TATTGCGTCT CCTGGAAGTT
401 CGATAACTAC CGGCAAAGGA GCCGTTAGCT GCTCTACGGG TAGCTTGAGT
25 451 TTGACAAAAA TGTCAGTTTG CTCTTCAGCA AAAACTTTTC AACCGATAAT
501 GGCGGTGCTA TCACGCAAAA AACTCTTTCA TTAA
```

This protein is frame-shifted with respect to cp0015.

The PSORT algorithm predicts an inner membrane location (0.047).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 9A. A

- 30 GST-fusion was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in an immunoassay (Figure 9B) and for FACS analysis (Figure 9C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 35 These experiments suggest that cp0014 is a useful immunogen. These properties are not evident from the sequence alone.

Example 10

The following *C.pneumoniae* protein (PID 6172317) was expressed <SEQ ID 19; cp0015>:

```

1  MSALFSENTS SKKGAIQTS DALTITGNQG EVSFSNTSS DSGAAIFTEA
51  SVTISNNAKV SFIDNKVTGA SSSTTGDMG GAICAYKTST DTKVTLTGNQ
40 101 MLLFSNNTST TAGGAIYVKK LELASGGLTL FSRNSVNGGT APKGGAIAIE
151 DSGELSLSAD SGDIVFLGNT VTSTTPGTNR SSIDLGTSK MTALRSAAGR
```

5
10

```

201 AIYFYDPITT GSSTTVTDVL KVNETPADSA LQYTGNIIFT GEKLSETEAA
251 DSKNLTSKLL QPVTLSGGTL SLKHGVTLQT QAFYQQADSR LEMDVGTGLE
301 PADTSTINNL VINISSIDGA KKAKIETKAT SKNLTLSGTI TLDDPTGTFY
351 ENHSLRNPQS YDILELKASG TVTSTAVTFD PIMGEKFHYG YQGTWGPPIVW
401 GTGASTTATF NWTKTGYIPN PERIGSLVPN SLWNAFIDIS SLHYLMETAN
451 EGLQGDRAFW CAGLSNFFHK DSTKTRRGFR HLSGGYVIGG NLHTCSDKIL
501 SAAFCQLFGR DRDYFVAKNQ GTVYGGTLYY QHNETYISLP CKLRPCSLSY
551 VPTEIPVLFS GNLSYTHTDN DLKTKYTTFP TVKGSWGNDS FALEFGGRAP
601 ICLDESALFE QYMPFMKLQF VYAHQEGFKE QGTEAREFGS SRLVNLALPI
651 GIRFDKESDC QDATYNLTG YTVDLVRSNP DCTTTLRISG DSWKTFGNL
701 ARQALVLRAG NHFCFNSNFE AFSQFSFELR GSSRNYNVDL GAKYQF*

```

This sequence is frame-shifted with respect to cp0014.

The cp0015 nucleotide sequence <SEQ ID 20> is:

15
20
25
30
35
40
45
50
55

```

1 ATGTCAGCTC TGTTTTCTGA AAATACCTCC TCAAAGAAAG GCGGAGCCAT
51 TCAGACTTCC GATGCCCTTA CCATTACTGG AAACCAAGGG GAAGTCTCTT
101 TTTCTGACAA TACTTCTTCG GATTCTGGAG CTGCAATTTT TACAGAAGCC
151 TCGGTGACTA TTTCTAATAA TGCTAAAGTT TCCTTTATTTG ACAATAAGGT
201 CACAGGAGCG AGCTCCTCAA CAACGGGGGA TATGTCAGGA GGTGCTATCT
251 GTGCTTATAA AACTAGTACA GATACTAAGG TCACCCTCAC TGGAAATCAG
301 ATGTTACTCT TCAGCAACAA TACATCGACA ACAGCGGGAG GAGCTATCTA
351 TGTGAAAAG CTGCAACTGG CTTCGGGAGG ACTTACCCTA TTCAGTAAAG
401 ATAGTGTCAA TGGAGGTACA GCTCCTAAAG GTGGAGCCAT AGCTATCGAA
451 GATAGTGGGG AATTGAGTTT ATCCGCCGAT AGTGGTGACA TTGTCTTTT
501 AGGGAATACA GTCACCTCTA CTACTCCTGG GACGAATAGA AGTAGTATCG
551 ACTTAGGAAC GAGTGCAAAG ATGACAGCTT TGCCTTCTGC TGCTGGTAGA
601 GCCATCTACT TCTATGATCC CATAACTACA GGATCATCCA CAACAGTTAC
651 AGATGTCTTA AAAGTTAATG AGACTCCGGC AGATTCTGCA CTACAATATA
701 CAGGGAACAT CATCTTCACA GGAGAAAAGT TATCAGAGAC AGAGGCCCGCA
751 GATTCTAAAA ATCTTACTTC GAAGCTACTA CAGCCTGTAA CTCTTTCAGG
801 AGGTACTCTA TCTTTAAAC ATGGAGTGAC TCTGCAGACT CAGGCATTCA
851 CTCAACAGGC AGATTCTCGT CTCGAAATGG ACGTAGGAAC TACTCTAGAA
901 CCTGCTGATA CTAGCACCAT AAACAATTG GTCATTAACA TCAGTTCTAT
951 AGACGGTGCA AAGAAGGCAA AAATAGAAAC CAAAGCTACG TCAAAAAATC
1001 TGACTTTATC TGGAAACATC ACTTTATTGG ACCCGACGGG CACGTTTTTAT
1051 GAAAATCATA GTTTAAGAAA TCCTCAGTCC TACGACATCT TAGAGCTCAA
1101 AGCTTCTGGA ACTGTAACAA GCACCGCAGT GACTCCAGAT CCTATAATGG
1151 GTGAGAAATT CCATTACGGC TATCAGGGAA CTGGGGCCC AATTGTTTGG
1201 GGGACAGGGG CTCTACGAC TGCAACCTTC AACTGGACTA AAAGTGGCTA
1251 TATTCTTAAT CCGGAGCGTA TCGGCTCTTT AGTCCCTAAT AGCTTATGCA
1301 ATGCATTTAT AGATATTAGC TCTCTCCATT ATCTTATGGA GACTGCAAAC
1351 GAAGGGTTGC AGGGAGACCG TGCTTTTGG TGTGCTGGAT TATCTAAGTT
1401 CTTCCATAAG GATAGTACAA AAACACGACG CGGGTTTCGC CATTTGAGTG
1451 GCGGTTATGT CATAGGAGGA AACCTACATA CTGTGTCAGA TAAGATTCTT
1501 AGTGCTGCAT TTTGTCAGCT CTTTGAAGA GATAGAGACT ACTTTGTAGC
1551 TAAGAAATCAA GGTACAGTCT ACGGAGGAAC TCTCTATTAC CAGCACACG
1601 AAACCTATAT CTCTCTTCTT TGCAAACTAC GGCCTTGTTC GTTGTCTTAT
1651 GTTCTTACAG AGATTCTCTG TCTCTTTTCA GGAAACCTTA GCTACACCCA
1701 TACGGATAAC GATCTGAAAA CCAAGTATAC AACATATCCT ACTGTTAAAG
1751 GAAGCTGGGG GAATGATAGT TTCGCTTTAG AATTCGGTGG AAGAGCTCCG
1801 ATTTGCTTAG ATGAAAAGTGC TCTATTGAG CAGTACATGC CCTTCATGAA
1851 ATTGCAGTTT GTCTATGCAC ATCAGGAAGG TTTTAAAGAA CAGGGAACAG
1901 AAGCTCGTGA ATTTGGAAGT AGCCGCTTGG TGAATCTTGC CTTACCTATC
1951 GGGATCCGAT TTGATAAGGA ATCAGACTGC CAAGATGCAA CGTACAATCT
2001 AACTCTTGGT TATACTGTGG ATCTTGTTCG TAGTAACCCC GACTGTACGA
2051 CAACACTGCG AATTAGCGGT GATTCTTGGG AAACCTTCGG TACGAATTGG
2101 GCAAGACAAG CTTTAGTCTT TCGTGACGGG AACCATTTTT GCTTTAATCT
2151 AAATTTTGAA GCCTTTAGCC AATTTCTTTT TGAATTGCGT GGGTCATCTC
2201 GCAATTACAA TGTAGACTTA GGAGCAAAAT ACCAATCTTA A

```

The PSORT algorithm predicts a cytoplasmic location (0.274).

60 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 10A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 10B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp0015 is a useful immunogen. These properties are not evident from the sequence alone.

Example 11

The following *C.pneumoniae* protein (PID 6172325) was expressed <SEQ ID 21; cp0019>:

```

5      1  LQDSQDYSFV  KLSFGAGGTI  ITQDASQKPL  EVAPSRPHYG  YQGHWNVQVI
      51  PGTGTQPSQA  NLEWVRTGYL  PNPERQGSIV  PNSLWGSFVD  QRAIQEIMVN
     101  SSQILCQERG  VWGAGIANFL  HRDKINEHGY  RHSGVGYLVG  VGTHAFSDAT
     151  INAAFQQLFS  RDKDYVVSKN  HGTSYSGVVF  LEDTLEFRSP  QGFYTDSSSE
     201  ACCNQVVTID  MQLSYSHRNN  DMKTKYTTYT  EAQGSWANDV  FGLEFGATTY
    10  251  YYPNSTFLFD  YYSFPLRLQC  TYAHQEDFKE  TGGEVRHFTS  GDLFNLAVPI
     301  GVKFERFSDC  KRGSYELTLA  YVPDVIRKDP  KSTATLASGA  TWSTHGNLNS
     351  RQGLQLRLGN  HCLINPGIEV  FSHGATELRG  SSRNYNINLG  GKRYR*

```

This sequence is frame-shifted with respect to cp0018.

The cp0019 nucleotide sequence <SEQ ID 22> is:

```

15      1  TTGCAAGACT  CTCAAGACTA  TAGCTTTGTA  AAGTTATCTC  CAGGAGCGGG
      51  AGGGACTATA  ATTACTCAAG  ATGCTTCTCA  GAAGCCTCTT  GAAGTAGCTC
     101  CTTCTAGACC  ACATTATGGC  TATCAAGGAC  ATTGGAATGT  GCAAGTCATC
     151  CCAGGAACGG  GAACTCAACC  GAGCCAGGCA  AATTAGAAAT  GGGTGCAGAC
     201  AGGATACCTT  CCGAATCCCG  AACGGCAAGG  ATCTTTAGTT  CCCAATAGCC
     251  TGTGGGGTTC  TTTTGTGAT  CAGCGTGCTA  TCCAAGAAAT  CATGTTAAAT
     301  AGTAGCCAAA  TCTTATGTCA  GGAACGGGGA  GTCTGGGGAG  CTGGAATTGC
     351  TAATTTCCTA  CATAGAGATA  AAATTAATGA  GCACGGCTAT  CGCCATAGCG
     401  GTGTCGGTTA  TCTTGTGGGA  GTTGGCACTC  ATGCTTTTTC  TGATGCTACG
     451  ATAAATGCGG  CTTTTTGCCA  GCTCTTCAGT  AGAGATAAAG  ACTACGTAGT
    25  501  ATCCAAAAAT  CATGGAACCT  GCTACTCAGG  GGTCTGATTT  CTTGAGGATA
     551  CCCTAGAGTT  TAGAAGTCCA  CAGGGATTCT  ATACTGATAG  CTCCTCAGAA
     601  GCTTGCTGTA  ACCAAGTCGT  CACTATAGAT  ATGCAGTTGT  CTTACAGCCA
     651  TAGAAATAAT  GATATGAAAA  CCAAATACAC  GACATATCCA  GAAGCTCAGG
     701  GATCTTGGGC  AAATGATGTT  TTTGGTCTTG  AGTTTGGAGC  GACTACATAC
    30  751  TACTACCCTA  ACAGTACTTT  TTTATTGAT  TACTACTCTC  CGTTTCTCAG
     801  GCTGCAGTGC  ACCTATGCTC  ACCAGGAAGA  CTTCAAAGAG  ACAGGAGGTG
     851  AGGTTGCTCA  CTTTACTAGC  GGAGATCTTT  TCAATTTAGC  AGTTCCTATT
     901  GGCCTGAAGT  TTGAGAGATT  TTCAGACTGT  AAAAGGGGAT  CTTATGAAC
     951  TACCCTTGCT  TATGTTCCTG  ATGTGATTTC  CAAAGATCCC  AAGAGCACGG
    35 1001  CAACATTGGC  TAGTGGAGCT  ACGTGGAGCA  CCCACGGAAA  CAATCTCTCC
     1051  AGACAAGGAT  TACAAGTCGG  TTTAGGGAAC  CACTGTCTCA  TAAATCCTGG
     1101  AATTGAGGTG  TTCAGTCACG  GAGCTATTGA  ATTGCGGGGA  TCCTCTCGTA
     1151  ATTATAACAT  CAATCTCGGG  GGTAAATACC  GATTTTAA

```

The PSORT algorithm predicts a cytoplasmic location (0.189).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 11A. This protein was used to immunise mice, whose sera were used in a Western blot (Figure 11B) and an immunoblot assay (Figure 11C). A his-tagged protein was also expressed.

These experiments show that cp0019 is a useful immunogen. These properties are not evident from the sequence alone.

45 Example 12

The following *C.pneumoniae* protein (PID 4376466) was expressed <SEQ ID 23; cp6466>:

```

      1  MRKISVGICI  TILLSLSVVL  QGCKESSHSS  TSRGELAINI  RDEPRSLDPR
     51  QVRLLEISL  VKHIYEGLVQ  ENNLSGNIEP  ALAEDYSLSS  DGLTYTFKLK
    101  SAFWSNGDPL  TAEDFIESWK  QVATQEVSGI  YAFALNPIKN  VRKIQEGHLS
    151  IDHFGVHSPN  ESTLVVTLES  PTHFLKLLA  LPVFFPVHKS  QRTLQSKSLP
    201  IASGAFYPKN  IKQKQWIKLS  KNPHYYNQSQ  VETKTTIHF  IPDANTAACL

```

5 251 FNQKLNWQG PPWGERIPQE TLSNLQSKGH LHSFDVAGTS WLTFNINKFP
 301 LNNMKLREAL ASALDKREALV STIFLGRAKT ADHLLPTNIH SYPEHQKQEM
 351 AQRQAYAKKL FKEALEELQI TAKDLEHLNL IFPVSSSASS LLVQLIREQW
 401 KESLGFAIPI VGKEFALLQA DLSSGNFSLA TGGWFADFAD PMAFLTIFAY
 451 PSGVPPYAIN HKDFLEILQN IEQEQDHQKR SELVSQASLY LETFHTIIEPI
 501 YHDAFQFAMN KKLSNLGVSP TGVVDFRYAK EN*

A predicted signal peptide is highlighted.

The cp6466 nucleotide sequence <SEQ ID 24> is:

10 1 ATGCGCAAGA TATCAGTGGG AATCTGTATC ACCATTCTCC TTAGCCTCTC
 51 CGTAGTCCTC CAAGGCTGCA AGGAGTCCAG TCACTCCTCT ACATCTCGGG
 101 GAGAACTCGC TATTAATATA AGAGATGAAC CCCGTTCTTT AGATCCAAGA
 151 CAAGTGCGAC TTCTTTCAGA AATCAGCCTT GTCAAACATA TCTATGAGGG
 201 ATTAGTTCAG AAAAATAATC TTTCAGGAAA TATAGAGCCT GCTCTTGCAG
 15 251 AAGACTACTC TCTTTCCTCG GACGGACTCA CTTATACTTT TAAACTGAAA
 301 TCAGCTTTTT GGAGTAATGG CGACCCCTTA ACAGCTGAAG ACTTTATAGA
 351 ATCTTGGAAA CAAGTAGCTA CTCAAGAACT CTAGGAATC TATGCTTTTG
 401 CCTTGAATCC AATTAAAAAT GTACGAAAGA TCCAAGAGGG ACACCTCTCC
 451 ATAGACCATT TTGGAGTGCA CTCTCCTAAT GAATCTACAC TTGTTGTTAC
 501 CCTGGAATCC CCAACCTCGC ATTCTTAA AACTTTAGCT CTTCAGTCT
 20 551 TTTTCCCCGT TCATAAATCT CAAAGAACCC TGCAATCCAA ATCTCTACCT
 601 ATAGCAAGCG GAGCTTTCTA TCCTAAAAAT ATCAAACAAA AACAAATGGAT
 651 AAAACTCTCA AAAAACCTC ACTACTATAA TCAAAGTCAG GTGGAACTA
 701 AAACGATTAC GATTCACTTC ATTCCCGATG CAAACACAGC AGCAAACTA
 15 751 TTTAATCAGG GAAAACCTCA TTGGCAAGGA CCTCCTTGGG GAGAACGCAT
 25 801 TCCTCAAGAA ACCCTATCCA ATTTACAGTC TAAGGGGCAC TTACACTCTT
 851 TTGATGTCGC AGGAACCTCA TGGCTCACCT TCAATATCAA TAAATTCCCC
 901 CTCACAAATA TGAAGCTTAG AGAAGCCTTA GCATCAGCCT TAGATAAGGA
 951 AGCTCTTGTC TCAACTATAT TCTTAGGCCG TGCAAAAACT GCCGATCATC
 1001 TCCTACCTAC AAATATTATC AGCTATCCCG AACATCAAAA ACAAGAGATG
 30 1051 GCACAACGCC AAGCTTACGC TAAAAAATC TTTAAAGAAG CTTAGAAGA
 1101 ACTCCAAATC ACTGCTAAAG ATCTCGAACA TCTTAATCTT ATCTTTCCCG
 1151 TTTCTCTGTC AGCAAGTTCT TTTACTAGTCC AACTTATACG AGAACAGTGG
 1201 AAAGAAAGTT TAGGGTTTCGC TATCCCTATT GTCGGAAAGG AATTGCTCT
 1251 TCTCCAAGCA GACCTATCTT CAGGGAACCT CTCTTTAGCT ACAGGAGGAT
 35 1301 GGTTCGCAGA CTTTGCTGAT CCTATGGCAT TTCTAACGAT CTTTGCTTAT
 1351 CCATCAGGAG TTCCTCCTTA TGCAATCAAC CATAAGGACT TCCTAGAAAT
 1401 TCTACAAAC ATAGAACAAG AGCAAGATCA CCAAAAACGC TCGGAATTAG
 1451 TGTCGCAAGC TTCTCTTATC CTAGAGACCT TTCAATATAT TGAGCCGATC
 1501 TACCACGAGC CATTTCAATT TGCTATGAAT AAAAACTTT CTAATCTAG
 40 1551 AGTCTACCA ACAGGAGTTG TGGACTTCCG TTATGCTAAG GAAATTAG

The PSORT algorithm predicts that the protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified both as a GST-fusion product and a His-tag fusion product. Purification of the protein as a GST-fusion product is shown in Figure 12A. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 12B and 12C). FACS analysis was also performed.

These experiments show that cp6466 is a useful immunogen. These properties are not evident from the sequence alone.

Example 13

The following *C.pneumoniae* protein (PID 4376468) was expressed <SEQ ID 25; cp6468>:

50 1 MFSRWITLFL LFISLTGCS YSSKHQSLI IPIHDDPVAF SPEQAKRAMD
 51 LSIAQLLFDG LTRETHRESN DLELAIASRY TVSEDFCSYT FFIKDSALWS
 101 DGTPITSEDI RNAWEYAQEN SPHIQIFQGL NFSTPSSNAI TIHLDSPNPD
 151 FPKLLAFPAF AIFKPENPKL FSGPYTLVEY FPGHNIHLKK NPNYDYHCV
 201 SINSIKLLII PDYTAIHL NRGKVDWVGQ PWHQGI PWEL HKQSQYHYT
 55 251 YPVEGAFWLC LNTKSPHLND LQNRHRLATC IDKRSTIEEA LQGTQQPAET

301 LSRGAPQPNQ YKKQKPLTPQ EKLVLTYPSD ILRCQRIAEI LKEQWKAAGI
 351 DLILEGLEHY LFNKRRKVD YAIATQTGVA YYPGANLISE EDKLLQNFEI
 401 IPIYYLSYDY LTQDFIEGVI YNAGAVDLK YTYFP*

A predicted signal peptide is highlighted.

5 The cp6468 nucleotide sequence <SEQ ID 26> is:

1 ATGTTTTTAC GATGGATCAC CCTCTTTTTA TTATTCATTA GCCTTACTGG
 51 ATGCTCCTCC TACTCTTCAA AACATAAACA ATCTTTAATT ATTCCCATAC
 101 ATGACGACCC TGTAGCTTTT TCTCCTGAAC AAGCAAAACG GGCCATGGAC
 151 CTTTCTATTG CCCAATTCT TTTTGATGGT CTGACTAGAG AAACATCATC
 10 201 CGAATCCAAT GATTTGGAAT TAGCGATTGC CAGTCGCTAT ACAGTCTCTG
 251 AAGACTTTTG CTCTTATACG TTCTTTATCA AAGACAGCGC TTTATGGAGC
 301 GACGGAACAC CAATCACCTC CGAAGATATC CGTAACGCTT GGGAGTATGC
 351 ACAGGAGAAC TCTCCCCACA TACAGATCTT CCAAGGACTT AACTTCTCAA
 401 CTCCTTCATC AAATGCAATT ACGATTATC TCGACTCGCC CAACCCCGAT
 15 451 TTTCCTAAGC TTCTTGCTT TCCTGCATTT GCTATCTTTA AACCAGAAAA
 501 CCCGAAGCTC TTTAGCGGTC CGTATACTCT TGTAGAGTAT TTCCCAGGGC
 551 ATAACATTTCA TTAAAGAAA AACCTAATC ATTACGACTA CCACTGCGTC
 601 TCCATCAACT CCATCAAACT GCTCATTATT CCTGATATAT ATACAGCCAT
 651 CCACCTCCTA AACAGAGGCA AGGTGGACTG GGTAGGACAA CCCTGGCATC
 20 701 AAGGGATTCC TTGGGAGCTC CATAAACAAT CGCAATATCA CTACTACACC
 751 TATCCTGTAG AAGGTGCTTT CTGGCTTTGT CTAATACAA AATCCCCACA
 801 CTTAAATGAT CTTCAAAACA GACATAGACT CGCTACTTGT ATTGATAAAC
 851 GTTCTATCAT TGAAGAAGCT CTTCAAGGAA CCCAACAACC AGCGGAAACA
 901 CTGTCCCGAG GAGCTCCACA ACCAAATCAA TATAAAAAAC AAAAGCCTCT
 25 951 AACTCCACAA GAAAACTCG TGCTTACCTA TCCCTCAGAT ATTCTAAGAT
 1001 GCCAACGCAT AGCAGAAATC TTAAGGAAC AATGAAAGC TGCTGGAATA
 1051 GATTTAATCC TTGAAGGACT CGAATACCAT CTGTTTGTTA ACAACGAAA
 1101 AGTCCAAGAC TACGCCATAG CAACACAGAC TGGAGTTGCT TATTACCCAG
 1151 GAGCAATCT AATTCTGAA GAAGACAAGC TCCTGCAAAA CTTTGAGATT
 30 1201 ATCCCGATCT ACTATCTGAG CTATGACTAT CTCACTCAAG ATTTTATAGA
 1251 GGGAGTAATC TATAATGCTT CTGGAGCTGT AGATCTCAA TATACCTATT
 1301 TCCCCTAG

The PSORT algorithm predicts that this protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 13A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 13B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6468 is a useful immunogen. These properties are not evident from the sequence alone.

Example 14

40 The following *C.pneumoniae* protein (PID 4376469) was expressed <SEQ ID 27; cp6469>:

1 MKMHLKPTL KSLIPNLLFL LLTLSSCSKQ KQELGKHLV IAMSHDLADL
 51 DFRNAYLSRD ASLAKALYEG LTRETDQGIA LALAESYTLS KDHKVYTFKL
 101 RPSVWSDGTP LTAYDFEKSI KQLYFEEFSP SIHTLLGVIK NSSAIHNAQK
 151 SLETLGIOAK DDLTLVITL QFPFYFLTLI ARPVFSPVHH TLRESYKKG
 45 201 PPSTYISNGP FVLKKEHQN YLILEKNPHY YDHESVKLDR VTLKIIPDAS
 251 TATKLFKSKS IDWIGSPWSA PISNEDQKVL SQEKILYTSV SSTTLIYNL
 301 QKPLIQNKAL RKALAHADR KSILRLVPSG QEAVTLVPPN LSQNLQKEI
 351 STEERQTKAR AYFQEAETL SEKELAEISI LYPIDSSNSS IIAQBIQRQL
 401 KDTLGLIKI QGMEYHCFK KRRQGDFFIA TGGWIAEYVS PVAFLSILGN
 50 451 PRDLTQWRNS DYKTFLEKLY LPHAYKENLK RAEMIEEET PIIPLYHGKY
 501 IYAIHPKIQN TFGSLLGHTD LKNIDILS*

A predicted signal peptide is highlighted.

The cp6469 nucleotide sequence <SEQ ID 28> is:


```

1  ATGAAGATGC ATAGGCTTAA ACCTACCTTA AAAAGTCTGA TCCCTAATCT
51  TCTTTTCTTA TTGCTCACTC TTTCAAGCTG CTCAAAGCAA AAACAAGAAC
101 CCTTAGGAAA ACATCTCGTT ATTGCGATGA GCCATGATCT CGCCGACCTA
151 GATCCTCGCA ATGCCTATTT AAGCAGAGAT GCTTCCCTAG CAAAAGCCCT
201 CTATGAAGGA CTGACAAGAG AAAGTATGCA AGGAATCGCA CTGGCTCTTG
251 CAGAAAGTTA TACCCTGTCA AAAGATCATA AGGTCTATAC CTTTAAACTC
301 AGACCTTCTG TGTGGAGCGA TGGCACTCCA CTCCTGCTT ATGACTTTGA
351 AAAATCTATA AAACAAGTGT ACTTCGAAGA ATTTTCACCT TCCATACATA
401 CTTTACTCGG CGTGATTTAA AATCTTTCGG CAATCCACAA TGCTCAAAAA
451 TCTCTGAAA CTCTTGGGAT ACAGGCAAAA GATGATCTTA CTTTGGTGAT
10 501 TACCCTAGAG CAACCTTTCC CATACTTTCT CACACTTATC GCTCGCCCCG
551 TATCTCTCCC TGTTCATCAC ACCCTTAGGG AATCCTATAA GAAAGGAACA
601 CCCCCATCCA CATACATCTC CAATGGGCCC TTTGTCTTAA AAAACATGA
651 ACACCAAAAC TACTTAATT TAGAAAAAA TCCTCACTAC TATGATCATG
15 701 AATCAGTAAA GTTAGACCGA GTCACCTTAA AAATATATCC AGACGCCTCC
751 ACAGCCACGA AACTTTTCAA AAGTAAATCT ATAGATTGGA TTGGCTCACC
801 TTGGAGCGCT CCGATATCTA ACGAAGACCA AAAAGTTCTC TCCCAAGAAA
851 AGATTCTTAC CTATTCTGTT TCAAGCACCA CCTTCTTAT CTATAACCTG
20 901 CAAAAACCTC TAATACAAA TAAAGCCCTC AGGAAAGCCA TTGCTCATGC
951 TATTGATAGA AAATCTATCT TAAGACTCGT GCCTTCAGGA CAAGAAGCTG
1001 TAACTCTAGT TCCCCAAAT CTTTCACAAC TCAATCTTCA AAAAGAGATC
1051 TCAACAGAAG AACGACAAC AAAAGCCAGA GCATATTTTC AAGAAGCTAA
1101 AGAAACACTT TCTGAAAAAG AACTCGCAGA ACTCAGCATC CTCTATCCTA
1151 TAGATTCTCT GAATTCCTCC ATCATAGCTC AAGAAATCCA AAGACAACCT
25 1201 AAAGATACCT TAGGATTGAA AATCAAAATC CAAGGCATGG AGTACCACCTG
1251 CTTTTTAAAG AAACGTCGTC AAGGAGATTT CTTTCATAGCG ACAGGAGGAT
1301 GGATTGCGGA ATACGTAAGC CCCGTAGCCT TCCTATCTAT TCTAGGCAAC
1351 CCCAGAGACC TCACACAATG GAGAAACAGT GATTACGAAA AGACTTTTGA
1401 GAAACTCTAT CTCCCTCATG CCTACAAAGA GAATTTAAAA CGCGCAGAAA
30 1451 TGATAATAGA AGAAGAAACC CCGATTATCC CCCTGTATCA CGGCAAAATAT
1501 ATTTACGCTA TACATCCTAA AATCCAGAAT ACATTCGGAT CTCCTCTAGG
1551 CCACACAGAT CTCAAAAATA TCGATATCTT AAGTTAG

```

The PSORT algorithm predicts a periplasmic location (0.934).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 14A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 14B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6469 is a useful immunogen. These properties are not evident from the sequence alone.

Example 15

40 The following *C.pneumoniae* protein (PID 4376602) was expressed <SEQ ID 29; cp6602>:

```

1  MAASGGTGGL GGTQGVNLAA VEAAAAKADA AEVVASQEGS EMNMIQQSQD
51  LTNFAAATRT KKKEEFQTL ESRKKGEAGK AEKKSESTEE KPDTDLADKY
101 ASGNSEISGQ ELRGLRDAIG DDASPEDILA LVQEKIKDPA LQSTALDYLV
151 QTTPPSQGKL KEALIQARNT HTEQFGRTAI GAKNILFASQ EYADQLNVSP
45 201 SGLRSLYLEV TGDTHTCDDL LSMLQDRYTY QDMAIVSSFL MKGMATELKR
251 QGPYVPSAQL QVLMTEFRNL QAVLTSYDYF ESRVPILDS LKABGIQTPS
301 DLNFVKVAES YHKIINDKFP TASKVEREVR NLIGDDVDSV TGVNLNLFSA
351 LRQTSSRLFS SADKRQQLGA MIANALDAVN INNEDYPKAS DFPKPYPWS*

```

The cp6602 nucleotide sequence <SEQ ID 30> is:

```

50 1  ATGGCAGCAT CAGGAGGCAC AGGTGGTTTA GGAGGCACTC AGGGTGTCAA
51  CCTTGACGCT GTAGAAGCTG CAGCTGCAAA AGCAGATGCA GCAGAAGTTG
101 TAGCCAGCCA AGAAGGTTCT GAGATGAACA TGATTCAACA ATCTCAGGAC
151 CTGACAAATC CCGCAGCAGC AACACGCACG AAAAAAAGG AAGAGAAGTT
201 TCAAACTCTA GAATCTCGGA AAAAAGGAGA AGCTGCAAG GCTGAGAAAA
55 251 AATCTGAATC TACAGAAGAG AAGCCTGACA CAGATCTTGC TGATAAGTAT
301 GCTTCTGGGA ATTCTGAAAT CTCTGGTCAA GAACCTCGCG GCCTGCGTGA
351 TGCAATAGGA GACGATGCTT CTCCAGAAGA CATCTTGTCT CTTGTACAAG

```

401 AGAAATTAAG AGACCCAGCT CTGCAATCCA CAGCTTTGGA CTACCTGGTT
 451 CAAACGACTC CACCCTCCCA AGGTAAATTA AAAGAAGCGC TTATCCAAGC
 501 AAGGAATACT CATACGGAGC AATTCGGAGC AACTGCTATT GGTGCGAAAA
 551 ACATCTTATT TGCCTCTCAA GAATATGCAG ACCAACTGAA TGTTTCTCCT
 601 TCAGGGCTTC GCTCTTTGTA CTAGAAGTG ACTGGAGACA CACATACCTG
 651 TGATCAGCTA CTTTCTATGC TTCAAGACCG CTATACCTAC CAAGATATGG
 701 CTATTGTCAG CTCCTTTCTA ATGAAAGGAA TGGCAACAGA ATTA AAAAGG
 751 CAGGGTCCCT ACGTACCCAG TCGCAACTA CAAGTTCTCA TGACAGAAAC
 801 TCGTAACCTG CAAGCAGTTC TTACCTCGTA CGATTACTTT GAAAGTCGCG
 851 TTCCTATTTT ACTCGATAGC TTAAGAGCTG AGGGAATCCA AACTCCTTCT
 901 GATCTAAACT TTGTGAAGGT AGCTGAGTCC TACCATAAAA TCATTACCGA
 951 TAAGTTCCCA ACAGCATCTA AAGTAGAACG AGAAGTCCGC AATCTCATAG
 1001 GAGACGATGT TGATTCTGTG ACCGGTGTCT TGAACCTATT CTTTTCTGCT
 1051 TTACGTCAAA CGTCGTACG CCTTTTCTCT TCAGCAGACA AACGTCAGCA
 1101 ATTAGGAGCT ATGATTGCTA ATGCTTTAGA TGCTGTAAAT ATAAACAATG
 1151 AAGATTATCC CAAAGCATCA GACTTCCCTA AACCCTATCC TTGGTCATGA

The PSORT algorithm predicts a cytoplasmic location (0.080).

The protein was expressed in *E. coli* and purified as both a His-tag and a GST-fusion product, as shown in Figure 15A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 15B) and for FACS analysis (Figure 15C).

The cp6602 protein was also identified in the 2D-PAGE experiment (Cpn0324).

These experiments show that cp6602 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 16

The following *C. pneumoniae* protein (PID 4376727) was expressed <SEQ ID 31; cp6727>:

1 **MYKSLFWLLT** SSALVFSLHP LMAANTDLSS SDNYENGSSG SAAFTAKETS
 51 DASGTTYTLT SDVSIITNVA ITPADKSCFT NTGGALSFVG ADHSLVLQTI
 101 ALTHDGAAIN NTNTALSPFG FSSLLIDSAP ATGTSQGGKA ICVTNTEGGT
 151 ATFTDNASVT LQKNTSEKDG AAVSAYSIDL AKTTTAAALLD QNTSTKNGGA
 201 LCSTANTTVQ GNSGTVTFSS NTATDKGGGI YSKEKSTLD ANTGVVTFKS
 251 NTAKTGGAWS SDDNLALTGN TQVLFQENKT TGSAAQANNP ECGGGAICCY
 301 LATATDKTGL AISQNQEMSF TSNTTTANGG AIYATKCTLD GNTTLTFDQD
 351 TATAGCGGAI YTEDEDFSLK GSTGTVTFST NTAKTGGALY SKGNSSLTGN
 401 TNLLFSGNKA TGPSNSSANQ EGCGGAILAF IDSGSVSDKT GLSIANNQEV
 451 SLTSNAATVS GGAIYATKCT LTGNGSLTFD GNTAGTSGGA IYTETEDFTL
 501 TGSTGTVTFSS TNTAKTGGAL YSKGNNSLSG NTNLLFSGNK ATGPSNSSAN
 551 QEGCGGAILS FLBSASVSTK KGLWIEDNEN VSLSGNTATV SGGAIYATKC
 601 ALHGNTTLTF DGNTAETAGG AIYTETEDFT LTGSTGTVTF STNTAKTAGA
 651 LHTKGNTSFT KNKALVFSGN SATATATTTT DQEGCGGAIL CNISESDIAT
 701 KSLTLTENES LSFINTAKR SGGGIYAPKC VISGSESINF DGNTAETSGG
 751 AIYSKNLSIT ANGVSFTNN SGGKGGAIYI ADSGELSLEA IDGDTFSGN
 801 RATEGTSTPN SIHLGAGAKI TKLAAAPGHT IYFYDPITME APASGGTIEE
 851 LVINPVVKAI VPPQPKNGP IASVPVVPVA PANPNTGTIV FSSGKLPSQD
 901 ASIPANTTTI LNQKINLAGG NVVLKEGATL QVYSFTQQPD STVFMDAGTT
 951 LETTTTNTD GSIDLKNSLV NLDALDGKRM ITIAVNSTSG GLKISGDLKF
 1001 HNNESGFYDN PGLKANLNL PFLDLSSTSGT VNLDDEFNPIP SSMAAPDYGY
 1051 QGSWTLVPKV GAGGKVTIVA EWQALGYTPK PELRATLVPN SLWNAVYNH
 1101 SIQGEIATAM SDAPSHPGIW IGGIGNAFHQ DKQKENAGFR LISRGYIVGG
 1151 SMTTPQEYTF AVAFSQLFGK SKDYVVS DIK SQVYAGSLCA QSSYVIPLHS
 1201 SLRRHVL SKV LPELPGETPL VLGQVSYGR NHHNMTTKLA NNTQGKSDWD
 1251 SHSFAVEVGG SLPVDLNYRY LYSYSPYVKL QVSVVNQKGF QEVAADPRIF
 1301 DASHLVNVS I PMGLTFKHES AKPPSALLLT LGYAVDAYRD HPHCLTSLTN
 1351 GTSWSTFATN LSRQAFFAEA SGHLKLLHGL DCFASGSCEL RSSRSYNAN
 1401 CGTRYSF*

A predicted signal peptide is highlighted.

The cp6727 nucleotide sequence <SEQ ID 32> is:

	1	ATGAAATATT	CTTTACCTTG	GCTACTTACC	TCTTCGGCTT	TAGTTTTC
	51	CCTACATCCA	CTAATGGCTG	CTAACACGGA	TCTCTCATCA	TCCGATAACT
5	101	ATGAAATGG	TAGTAGTGGT	AGCGCAGCAT	TCCTGCCAA	GGAAACTTCG
	151	GATGCTTCAG	GAACCTACCTA	CACTCTCACT	AGCGATGTTT	CTATTACGAA
	201	TGTATCTGCA	ATTACTCCTG	CAGATAAAG	CTGTTTAC	AACACAGGAG
	251	GAGCATTTAG	TTTGTGTGGA	GCTGATCACT	CATTGGTTCT	GCAAACCATA
	301	GCGCTTACGC	ATGATGGTGC	TGCAATTAAC	AATACCAACA	CAGCTCTTTC
10	351	TTTCTCAGGA	TTCTCGTCAC	TCTTAATCGA	CTCAGCTCCA	GCAACAGGAA
	401	CTTCGGGCGG	CAAGGTGCT	ATTTGTGTGA	CAAAATACAGA	GGGAGGTACT
	451	GCGACTTTTA	CTGACAATGC	CAGTGTCAAC	CTCCAAAAA	ATACTTCAGA
	501	AAAAGATGGA	GCTGCAGTTT	CTGCCTACAG	CATCGATCTT	GCTAAGACTA
	551	CGACAGCAGC	TCTCTTAGAT	CAAAATACTA	GCACAAAAA	TGGCGGGGCT
15	601	CTCTGTAGTA	CAGCAAAACAC	TACAGTCCAA	GGAAACTCAG	GAACGGTGAC
	651	CTTCTCCTCA	AATACTGCTA	CAGATAAAG	TGGGGGGATC	TACTCAAAAG
	701	AAAAGGATAG	CACGCTAGAT	GCCAAATACAG	GAGTCGTTAC	CTTCAAATCT
	751	AATACTGCAA	AGACGGGGGG	TGCTTGGAGC	TCTGATGACA	ATCTTGCTCT
	801	TACCGGCAAC	ACTCAAGTAC	TTTTTCAGGA	AAATAAAACA	ACCGGCTCAG
	851	CAGCACAGGC	AAATAACCCG	GAAGGTGTG	TGGGGGCAAT	CTGTTGTTAT
20	901	CTTGCTACAG	CAACAGACAA	AACTGGATTA	GCCATTTCTC	AGAATCAAGA
	951	AATGAGCTTC	ACTAGTAATA	CAACAACCTG	GAATGGTGGG	GCGATCTACG
	1001	CTACTAAATG	TACTCTGGAT	CAAAACACAA	CTCTTACCTT	CGATCAGAAT
	1051	ACTGCGACAG	CAGGATGTGG	CGGAGCTATC	TATACAGAAA	CTGAAGATTT
25	1101	TTCTCTTAAG	GGAAGTACGG	GAACCGTGAC	CTTCAGCACA	AATACAGCAA
	1151	AGACAGGCGG	CGCCTTATAT	TCTAAAGGAA	ACAGCTCGCT	GACTGGAAAT
	1201	ACCAACCTGC	TCTTTTCAGG	GAACAAGCT	ACGGGCCCCG	GTAATTTCTTC
	1251	AGCAAATCAA	GAGGGTTGCG	GTGGGGCAAT	CCTAGCCTTT	ATTGATTTCAG
	1301	GATCCGTAAG	CGATAAAACA	GGACTATCGA	TTGCAAAACA	CCAAGAAGTC
30	1351	AGCCTCACTA	GTAATGCTGC	AACAGTAAGT	GGTGGTGCGA	TCTATGCTAC
	1401	CAATGTACT	CTAATGGAA	ACGGCTCCCT	GACCTTTGAC	GGCAATACCTG
	1451	CTGGAACCTC	AGGAGGGGCG	ATCTATACAG	AACTGAAGA	TTTTACTCTT
	1501	ACAGGAAGTA	CAGGAACCGT	GACCTTCAGC	ACAAATACAG	CAAAGACAGG
	1551	CGGCGCCTTA	TATTCTAAAG	GCAACAACCT	TCTGTCTGGT	AATACCAACC
35	1601	TGCTCTTTTC	AGGGAACAAA	GCTACGGGCC	CGAGTAATTC	TTCAGCAAAT
	1651	CAAGAGGGTT	GCGGTGGGGC	AATCCTATCG	TTTCTTGAGT	CAGCATCTGT
	1701	AAGTACTAAA	AAAGGACTCT	GGATTGAAGA	TAACGAAAAC	GTGAGTCTCT
	1751	CTGGTAATAC	TGCAACAGTA	AGTGGCGGTG	CGATCTATGC	GACCAAGTGT
	1801	GCTCTGCATG	GAAACACGAC	TCTTACCTTT	GATGGCAATA	CTGCCGAAAC
40	1851	TGCAGGAGGA	CGGATCTATA	CAGAAACCGA	AGATTTTACT	CTTACGGGAA
	1901	GTACGGGAAC	CGTGACCTTC	AGCACAAATA	CAGCAAAGAC	AGCAGGGGCT
	1951	CTACATACTA	AAGGAAATAC	TTCTTTTACC	AAAAATAAGG	CTCTTGTTAT
	2001	TTCTGGAAAT	TCAGCAACAG	CAACAGCAAC	AACAACCTACA	GATCAAGAAG
	2051	GTGTGGGTGG	AGCGATCCTC	TGTAATATCT	CAGAGTCTGA	CATAGCTACA
45	2101	AAAAGCTTAA	CTCTTACTGA	AAATGAGAGT	TTAAGTTTCA	TTAACAATAC
	2151	GGCAAAAAGA	AGTGGTGGTG	GTATTTATGC	TCCTAAGTGT	GTAATCTCAG
	2201	GCACTGAATC	CATAAACTTT	GATGGCAATA	CTGCTGAAAC	TTCCGGGAGGA
	2251	GCGATTTATT	CGAAAAACCT	TTGATTACA	GCTAACGGTC	CTGTCTCCTT
	2301	TACCAATAAT	TCTGGAGGCA	AGGGAGGCGC	CATTTATATA	GCCGATAGCG
50	2351	GAGAACTTTC	CTTAGAGGCT	ATTGATGGGG	ATATTACTTT	CTCAGGGAAC
	2401	CGAGCGACTG	AGGGAACCTC	AACTCCCAAC	TCGATCCATT	TAGGTGCAGG
	2451	GGCTAAGATC	ACTAAGCTTG	CAGCAGCTCC	TGGTCATACG	ATTTATTTTT
	2501	ATGATCCTAT	TACGATGGAA	GCTCCTGCAT	CTGGAGGAAC	AATAGAGGAG
	2551	TTAGTCATCA	ATCCTGTTGT	CAAAGCTATT	GTTCCTCCTC	CCCAACCAAA
55	2601	AAATGGTCTT	ATAGCTTCAG	TGCCGTAGT	CCCTGTAGCA	CCTGCAAAAC
	2651	CAAAACACGG	AACATATAGTA	TTTTCTTCTG	GAAACTCCC	CAGTCAAGAT
	2701	GCCTCGATTC	CTGCAAAATAC	TACCACATA	CTGAACGAGA	AGATCAACTT
	2751	AGCAGGAGGA	AATGTCGTTT	TAAAGAAGG	AGCCACCCCTA	CAAGTATATT
	2801	CTTCACACA	GCAGCCTGAT	TCTACAGTAT	TCATGGATGC	AGGAACGACC
60	2851	TTAGAGACCA	CGACAACATA	CAATACAGAT	GGCAGCATCG	ATCTAAAGAA
	2901	TCTCTCTGTA	AATCTGGATG	CTTTAGATGG	CAAGCGTATG	ATAACGATTG
	2951	CCGTAAACAG	CACAAGTGGG	GGATTAAAAA	TCTCAGGGGA	TCTGAAATTC
	3001	CATAACAATG	AAGGAAGTTT	CTATGACAAT	CCTGGGTGTA	AAGCAAACCT
	3051	AAATCTTCCT	TTCTTAGATC	TTTCTTCTAC	TTCAGGAACCT	GTAATTTTAG
65	3101	ACGACTTCAA	TCCGATTCCCT	TCTAGCATGG	CTGCTCCGGA	TTATGGGTAT
	3151	CAAGGGAGTT	GGACTCTGGT	TCCTAAAGTA	GGAGCTGGAG	GGAGGTGAC
	3201	TTTGGTCGCG	GAATGGCAAG	CGTTAGGATA	CACTCCTAAA	CCAGAGCTTC
	3251	GTGCGACTTT	AGTTCCTAAT	AGCCTTTGGA	ATGCTTATGT	AAACATCCAT

5 3301 TCTATACAGC AGGAGATCGC CACTGCGATG TCGGACGCTC CCTCACATCC
 3351 AGGGATTGG ATTGGAGGTA TTGGCAACGC CTTCCATCAA GACAAGCAAA
 3401 AGGAAAATGC AGGATTCCGT TTGATTCCA GAGGTTATAT TGTTGGTGGC
 3451 AGCATGACCA CCCCTCAAGA ATATACCTTT GCTGTTGCAT TCAGCCAACT
 3501 CTTTGGCAAA TCTAAGGATT ACGTAGTCTC GGATATTAAA TCTCAAGTCT
 3551 ATGCAGGATC TCTCTGTGCT CAGAGCTCTT ATGTCATTCC CTTGCATAGC
 3601 TCATTACGTC GCCACGTCCT CTCTAAGGTC CTTCCAGAGC TCCAGGAGA
 3651 AACTCCCTTT GTTCTCCATG GTCAAGTTTC CTATGGAAGA AACCACCATA
 3701 ATATGACGAC AAAGCTTGCG AACAAACAC AAGGGAAATC AGACTGGGAC
 10 3751 AGCCATAGCT TCGCTGTGTA AGTCGGTGGT TCTCTTCTTG TAGATCTAAA
 3801 CTACAGATAC CTTACCAGCT ACTCTCCCTA TGTGAAACTC CAAGTTGTGA
 3851 GTGTAAATCA AAAAGGATTC CAAGAGTTG CTGCTGATCC ACGTATCTTT
 3901 GACGCTAGCC ATCTGGTCAA CGTGTCTATC CCTATGGGAC TCACCTTCAA
 3951 ACACGAATCA GCAAAGCCCC CCAGTGCTTT GCTTCTTACT TTAGGTTACG
 15 4001 CTGTAGATGC TTACCGGGAT CACCTCACT GCCTGACCTC CTTAACAAAT
 4051 GGCACCTCGT GGTCTACGTT TGCTACAAAC TTATCACGAC AAGCTTTCTT
 4101 TGCTGAGGCT TCTGGACATC TGAAGTTACT TCATGGTCTT GACTGCTTCG
 4151 CTTCTGGAAG TTGTGAACTG CGCAGCTCCT CAAGAAGCTA TAATGCAAAAC
 4201 TGTGGAATC GTTATTCTTT CTAA

20 The PSORT algorithm predicts an outer membrane location (0.915).

The protein was expressed in *E. coli* and purified as a his-tag product, as shown in Figure 16A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 16B) and for FACS analysis (Figure 16C). A GST-fusion protein was also expressed.

The cp6727 protein was also identified in the 2D-PAGE experiment (Cpn0444).

25 These experiments show that cp6727 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 17

The following *C. pneumoniae* protein (PID 4376731) was expressed <SEQ ID 33; cp6731>:

30 1 MKSSLHWFLI SSSLALPLSL NFSFAFAVVE INLGPTNSFS GPGTYTPPAQ
 51 TTNADGTIYN LTGDIVSITNA GSPTALTASC FKETTGNLSF QGHGYQFLLQ
 101 NIDAGANCTF TMTAANKLLS FSGFSYLSLI QTNATATGTG AIKSTGACSI
 151 QSNYSYFPGQ NFSNDNGGAL QGSSISLSLN PNLTFAKNKA TQKGGALYST
 201 GGITINNTLN SASFSENTAA NNGGAIYTEA SSFISNKAI SFINNSVTAT
 251 SATGGAIYCS STSAPKPVLT LSDNGELNFI GNTAITSGGA IYTDNLVLSS
 35 301 GGPTLFRNNS AIDTAAPLGG AIAIADSGSL SLSALGGDIT FEGNTVVKGA
 351 SSSQTTTRNS INIGMTNAKI VQLRASQNT IVFYDPITTS ITAALSDALN
 401 LINGPDLAGNP AYQGTIVFSG EKLSEAEAAE ADNLKSTIQQ PLTLAGGQLS
 451 LKSGVTLVAK SFSQSPGSTL LMDAGTTLET ADGITINNLV LNVDLSKETK
 501 KATLKATQAS QIVTLGSLSL LVDPSGNVYE DVSWNNPQVF SCLTLTADDP
 40 551 ANIHITDLAA DPLEKNPIHW GYQGNWALSW QEDTATSKA ATLITWKTGY
 601 NPNPERRGTL VANTLWGSFV DVRSIQQLVA TKVRQSQETR GIWCEGISNF
 651 FHKDSTKINK GFRHISAGYV VGATTLASD NLITAAFCQL FGKDRDHFIN
 701 KNRASAYAAS LHLQHLATLS SPSLLRYLPG SESEQPVLFQ AQISYIYSKN
 751 TMKTYTQAP KGESSWYNDG CALELASSLP HTALSHEGLF HAYFPFIKVE
 45 801 ASYIHQDSFK ERNTTLVRSF DSGDLINVSF PIGITFERFS RNERASYEAT
 851 VIYVADVYRK NPDCTTALLI NNTSWKTGT NLSRQAGIGR AGIFYAFSPN
 901 LEVTSNLSME IRGSSRSYNA DLGKRFQF*

A predicted signal peptide is highlighted.

The cp6731 nucleotide sequence <SEQ ID 34> is:

50 1 ATGAAATCCT CTCTTCATTG GTTTTAAATC TCGTCATCTT TAGCACTTCC
 51 CTGTGCTACTA AATTCTCTCG CGTTTGCTGC TGTGTGTGAA ATCAATCTAG
 101 GACCTACCAA TAGCTTCTCT GGACCAGGAA CCTACACTCC TCCAGCCCAA
 151 ACAACAAATG CAGATGGAAC TATCTATAAT CTAACAGGGG ATGTCTCAAT
 201 CACCAATGCA GGATCTCCGA CAGCTCTAAC CGCTTCTGCG TTAAAGAAA

251 CTACTGGGAA TCTTTCTTTC CAAGGCCACG GCTACCAATT TCTCCTACAA
 301 AATATCGATG CGGGAGCGAA CTGTACCTTT ACCAATACAG CTGCAAATAA
 351 GCTTCTCTCC TTTTCAGGAT TCTCCTATTT GTCACATAA CAAACCACGA
 401 ATGCTACCAC AGGAACAGGA GCCATCAAGT CCACAGGAGC TTGTTCTATT
 5 451 CAGTCGAAGT ATAGTTGCTA CTTTGGCCAA AACTTTTCTA ATGACAATGG
 501 AGGCGCCCTC CAAGGCAGCT CTATCAGTCT ATCGCTAAAC CCCAACCTAA
 551 CGTTTGCCAA AACAAAGCA ACGCAAAAAG GGGGTGCCCT CTATTCCACG
 601 GGAGGGATTA CAATTAACAA TACGTTAAAC TCAGCATCAT TTTCTGAAA
 651 TACCGCGGCG AACATGGCG GAGCCATTTA CACGGAAGCT AGCAGTTTTA
 10 701 TTAGCAGCAA CAAGCAATT AGCTTTATAA ACAATAGTGT GACCGCAACC
 751 TCAGCTACAG GGGGAGCCAT TTACTGTAGT AGTACATCAG CCCCCAAACC
 801 AGTCTTAAGT CTATCAGACA ACGGGGAAC TGAATTTATA GGAAATACAG
 851 CAATTACTAG TGGTGGGGCG ATTTATACTG ACAATCTAGT TCTTTCTTCT
 901 GGAGGACCTA CGCTTTTAA AACAACTCT GCTATAGATA CTGCAGCTCC
 15 951 CTTAGGAGGA GCAATTGCCA TTGCTGACTC TGGATCTTTG AGTCTTTCGG
 1001 CTCTTGGTGG AGACATCACT TTTGAAGGAA ACACAGTAGT CAAAGGAGCT
 1051 TCTTCGAGTC AGACCACTAC CAGAAATTCT ATTAACATCG GAAACACCAA
 1101 TGCTAAGATT GTACAGCTGC GAGCCTCTCA AGGCAATACT ATCTACTTCT
 20 1151 ATGATCCTAT AACAACTAGC ATCACTGCAG CTCCTCTAGA TGCTCTAAC
 1201 TTAAATGGTC CTGACCTTGC AGGGAATCCT GCATATCAAG GAACCATCGT
 1251 ATTTTCTGGA GAGAAGCTCT CGGAAGCAGA AGCTGCAGAA GCTGATAATC
 1301 TCAAATCTAC AATTCAAGCA CCTCTAATC TTGCGGGAGG GCAACTCTCT
 1351 CTTAAATCAG GAGTCACTCT AGTTGCTAAG TCCTTTTTCG AATCTCCGGG
 1401 CTCTACCTTC CTCATGGATG CAGGGACCAC ATTAGAAACC GCTGATGGGA
 25 1451 TCACTATCAA TAATCTTGT TCTCAATGTAG ATTCTTAAA AGAGACCAAG
 1501 AAGGCTACGC TAAAGCAAC ACAAGCAAGT CAGACAGTCA CTTTATCTGG
 1551 ATCGCTCTCT CTTGTAGATC CTTCTGGAAA TGTCTACGAA GATGCTCTCT
 1601 GGAATAACCC TCAAGTCTTT TCTTGTCTCA CTCCTACTGC TGACGACCCC
 1651 GCGAATATTC ACATCACAGA CTTAGCTGCT GATCCCCCTAG AAAAAATCC
 30 1701 TATCCATTGG GGATACCAAG GGAATTGGGC ATTATCTTGG CAAGAGGATA
 1751 CTGCGACTAA ATCAAAGCA GCGACTCTTA CCTGGACAAA AACAGGATAC
 1801 AATCCGAATC CTGAGCGTCG TGGAACTTAA GTTGCTAACA CGCTATGGGG
 1851 ATCCTTTGTT GATGTGCGCT CCATACAACA GCTTGAGGCC ACTAAAGTAC
 35 1901 GCCAATCTCA AGAACTCGC GGCATCTGGT GTGAAGGGAT CTCGAACCTC
 1951 TTCCATAAAG ATAGCACGAA GATAAATAAA GGTTTTCGCC ACATAAGTGC
 2001 AGGTTATGTT GTAGGAGCGA CTACAACATT AGCTTCTGAT AATCTTATCA
 2051 CTGCAGCCTT CTGCCAATTA TTCGGGAAAG ATAGAGATCA CTTTATAAAT
 2101 AAAAAATAGAG CTTCTGCCTA TGCAGCTTCT CTCCATCTCC AGCATCTAGC
 40 2151 GACCTTGTCT TCTCCAAGCT TGTACGCTA CCTTCTGGA TCTGAAAGTG
 2201 AGCAGCCTGT CCTCTTTGAT GCTCAGATCA GCTATATCTA TAGTAAAAAT
 2251 ACTATGAAAA CCTATTACAC CCAAGCACCA AAGGGAGAGA GCTCGTGGTA
 2301 TAATGACGGT TCGGCTCTGG AACTTGCGAG CTCCTACCA CACACTGCTT
 2351 TAAGCCATGA GGTCTCTTC CACGCGTATT TTCTTTTCAT CAAAGTAGAA
 45 2401 GCTTCGTACA TACACCAAGA TAGCTTCAA GAACGTAATA CTACCTTGGT
 2451 ACATCTTTC GATAGCGGTG ATTTAATTAA CGTCTCTGTG CTTATTGGAA
 2501 TTACCTTCGA GAGATTCTCG AGAAACGAGC GTGCGTCTTA CGAAGCTACT
 2551 GTCATCTACG TTGCCGATGT CTATCGTAAG AATCCTGACT GCACGACAGC
 2601 TCTCCTAATC AACAAATACCT CGTGGAAAAC TACAGGAACG AATCTCTCAA
 50 2651 GACAAGCTGG TATCGGAAGA GCAGGGATCT TTTATGCCTT CTCTCCAAAT
 2701 CTTGAGGTCA CAAGTAACCT ATCTATGAA ATTCGTGGAT CTTACAGCAG
 2751 CTACAATGCA GATCTTGAG GTAAGTTCCA GTTCTAA

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E. coli* and purified as a his-tag product, as shown in Figure 17A. A
 GST-fusion protein was also expressed. The recombinant proteins were used to immunise mice,
 55 whose sera were used in a Western blot (Figure 17B; his-tag) and for FACS analysis (Figure 17C;
 his-tag and GST-fusion).

The GST-fusion protein also showed good cross-reactivity with human sera, including sera from
 patients with pneumonitis. Less cross-reactivity was seen with the his-fusion.

These experiments show that cp6731 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 18

The following *C.pneumoniae* protein (PID 4376737) was expressed <SEQ ID 35; cp6737>:

```

5      1  MPLSFKSSSF  CLLACLCSAS  CAFAETRLGG  NFVPPITNQG  EEILLTSDFFV
      51  CSNFLGASFS  SSFINSSSNL  SLGKGLSLT  FTSCQAPTNS  NYALLSAAET
     101  LTFKNFSSIN  FTGNQSTGLG  GLIYGKDIVF  QSIKDLIFTT  NRVAYSPASV
     151  TTSATPAITT  VTTGASALQP  TDSLTVENIS  QSIKFFGNLA  NFGSAISSSP
     201  TAVVKFINNT  ATMSFSHNFT  SSGGGVIYGG  SLLLFENNSG  CIIFTANSCV
     251  NSLKGVTSS  GTYALGSGGA  ICPTGTPEL  KNNQKCTFS  YNGTPNDAGA
     301  IYAETCNIVG  NQGALLDSN  TAARNGGAI  AKVLNIQGRG  PIFSRNRAE
     351  KGAIFIGPS  VGDPKQST  LTLASEGDI  AFQGNMLNTK  PGIRNAITVE
     401  AGGEIVLSA  QGGSRLLVFD  PITHSLPTS  PSNKDITINA  NGASGSVVFT
     451  SKGLSSTELL  LPANTTTILL  GTVKIASGEL  KITDNVAVNV  LGFATQSGSQ
     501  LTLGSGGTLG  LATPTGAPAA  VDFITIGLAF  DPFSFLKRDF  VSASVNAAGT
     551  NVTLTGALVL  DEHDVTDLYD  MVSLOTPEAI  PIAVFKGATV  TKTGFDPGEI
     601  ATPSHYGYQG  KWSYTWSRPL  LIPAPDGGFP  GGPSPSANTL  YAVWNSDTLV
     651  RSTYILDPER  YGEIVSNLW  ISFLGNQAFS  DILQDVLLID  HPGLSITAKA
     701  LGAYVEHTPR  QGHEGFSGRY  GGYQAALSMN  YTDHTTLGLS  FGQLYGKTNA
     751  NPYDSRCSEQ  MYLLSFFGQF  PIVTQKSEAL  ISWKAAYGYS  KNHLNNTYLR
     801  PDKAPKSQGG  WHNNSYVLI  SAEHPFLNWC  LLTRPLAQAW  DLSGFISAEF
     851  LGGWQSKFTE  TGDQLRSFSR  GKGYNVSLPI  GCSSQWPF  KKAPSTLTIK
     901  LAYKPDYRV  NPHNIVTVVS  NQESTSISGA  NLRRHGLFVQ  IHDVVDLTED
     951  TQAFNLNYTFD  GKNGFTNHRV  STGLKSTF*

```

25 A predicted signal peptide is highlighted.

The cp6737 nucleotide sequence <SEQ ID 36> is:

```

      1  ATGCCTCTTT  CTTTCAAATC  TTCATCTTTT  TGTCTACTTG  CCTGTTTATG
     51  TAGTGCAAGT  TCGCGCTTTG  CTGAGACTAG  ACTCGGAGGG  AACTTTGTTC
    101  CTCCAATTAC  GAATCAGGGT  GAAGAGATCT  TACTCACTTC  AGATTTTGTT
    151  TGTTCAAACT  TCTTGGGGGC  GAGTTTTTC  AGTTCCTTTA  TCAATAGTTC
    201  CAGCAATCTC  TCCTTATTAG  GGAAGGGCCT  TTCCTTAACG  TTTACCTCTT
    251  GTCAAGCTCC  TACAAATAGT  AACTATGCGC  TACTTTCTGC  CGCAGAGACT
    301  CTGACCTTCA  AGAATTTTTC  TTCTATAAAC  TTTACAGGGA  ACCAATCGAC
    351  AGGACTTGGC  GGCCTCATCT  ACGGAAAAGA  TATTGTTTTC  CAATCTATCA
    401  AAGATTGAT  CTTCACTACG  AACCCTGTTG  CCTATTCTCC  AGCATCTGTA
    451  ACTACGTCGG  CAACTCCCGC  AATCACTACA  GTAACACAG  GAGCCTCTGC
    501  TCTCCAACCT  ACAGACTCAC  TCACTGTCGA  AAACATATCC  CAATCGATCA
    551  AGTTTCTTGG  GAACCTTGCC  AACTTCGGCT  CTGCAATTAG  CAGTTCTCCC
    601  ACGGCAGTCG  TTAATTCAT  CAATAACACC  GCTACCATGA  GCTTCTCCCA
    651  TAACTTTACT  TCGTCAGGAG  GCGGCGTGAT  TTATGGAGGA  AGCTCTCTCC
    701  TTTTGTGAAA  CAATTCTGGA  TGCATCATCT  TCACCGCCAA  CTCCTGTGTG
    751  AACAGCTTAA  AAGGCGTCAC  CCCTTCATCA  GGAACCTATG  CTTTAGGAAG
    801  TGGCGGAGCC  ATCTGCATCC  CTACGGGAAC  TTTTGAATTA  AAAACAATC
    851  AGGGGAAGTG  CACCTTCTCT  TATAATGGTA  CACCAAATGA  TCGGGGTGCG
    901  ATCTACGCCG  AAACCTGCAA  CATCGTAGGG  AACCAGGGTG  CCTTGCTCCT
    951  AGATAGCAAC  ACTGCAGCGA  GAAATGGCGG  AGCCATCTGT  GCTAAAGTGC
   1001  TCAATATTCA  AGGACGCGGT  CCTATTGAAT  TCTCTAGAAA  CCGCGCGGAG
   1051  AAGGTTGGAG  CTATTTTCAT  AGGCCCTCT  GTTGGAGACC  CTGCGAAGCA
   1101  AACATCGACA  CTTACGATTT  TGGCTTCCGA  AGGTGATATT  GCGTTTCCAA
   1151  GAAACATGCT  CAATACAAAA  CCTGGAATCC  GCAATGCCAT  CACTGTAGAA
   1201  GCAGGGGGAG  AGATTGTGTC  TCTATCTGCA  CAAGGAGGCT  CACGTCTTGT
   1251  ATTTTATGAT  CCCATTACAC  ATAGCCTCCC  AACCACAAGT  CCGTCTAATA
   1301  AAGACATTAC  AATCAACGCT  AATGGCGCTT  CAGGATCTGT  AGTCTTTACA
   1351  AGTAAGGGAC  TCTCCTCTAC  AGAACTCCTG  TTGCCTGCCA  ACACGACAAC
   1401  TATACTTCTA  GGAACAGTCA  AGATCGCTAG  TGGAGAAGTG  AAGATTACTG
   1451  ACAATGCGGT  TGTCATGTT  CTTGGCTTCG  CTACTCAGGG  CTCAGGTCAG
   1501  CTTACCTTGG  GCTCTGGAGG  AACCTTAGGG  CTGGCAACAC  CCACGGGAGC
   1551  ACCTGCCGCT  GTAGACTTTA  CGATTGGAAA  GTTAGCATTC  GATCCTTTTT
   1601  CCTTCCTAAA  AAGAGATTTT  GTTTCAGCAT  CAGTAAATGC  AGGCACAAAA
   1651  AACGTCACCT  TAACAGGAGC  TCTGGTTCTT  GATGAACATG  ACGTTACAGA

```

1701 TCTTTATGAT ATGGTGTGTCAT TACAAACTCC AGTAGCAATT CCTATCGCTG
 1751 TTTTCAAAGG AGCAACCGTT ACTAAGACAG GATTTCCTGA TGGGGAGATT
 1801 GCGACTCCAA GCCACTACGG CTACCAAGGA AAGTGGTCCT ACACATGGTC
 1851 CCGTCCCCCTG TTAATTCCAG CTCCTGATGG AGGATTTCCCT GGAGGTCCCT
 5 1901 CTCCTAGCGC AAATACTCTC TATGCTGTAT GGAATTCAGA CACTCTCGTG
 1951 CGTTCTACCT ATATCTTAGA TCCCGAGCGT TACGGAGAAA TTGTTCAGCAA
 2001 CAGCTTATGG ATTTCTTCTT TAGGAAATCA GGCATTCTCT GATATTCTCT
 2051 AAGATGTTCT TTTGATAGAT CATCCCGGGT TGTCCATAAC CGCGAAAGCT
 10 2101 TTAGGAGCCT ATGTCGAACA CACACCAAGA CAAGGACATG AGGCGCTTTTC
 2151 AGGTGCTAT GGAGGCTACC AAGCTGCGCT ATCTATGAAC TACACGGACC
 2201 ACACTACGTT AGGACTTTCT TTCGGGCAGC TTTATGGAAA AACTAACGCC
 2251 AACCCCTACG ATTCACGTG CTCAGAACAA ATGTATTTAC TCTCGTTCTT
 2301 TGGTCAATTC CCTATCGTGA CTCAAAAGAG CGAGGCCTTA ATTTCTTGA
 15 2351 AAGCAGCTTA TGGTTATTCC AAAAATCACC TAAATACCAC CTACCTCAGA
 2401 CCTGACAAAG CTCAAAATC TCAAGGGCAA TGGCATAACA ATAGTTACTA
 2451 TGTTCTTATT TCTGCAGAAC ATCCTTTCTT AAACCTGGTGT CTCTTACAA
 2501 GACCTCTGCG TCAAGCTTGG GATCTTTCAG GTTTTATTTT CGCAGAATTC
 2551 CTAGGTGGTT GGCAAGTAA GTTCACAGAA ACTGGAGATC TGCAACGCTAG
 2601 CTTTAGTAGA GGTAAAGGGT ACAATGTTTC CCTACCGATA GGATGTTCTT
 20 2651 CTCAATGGTT CACACCATTT AAGAAGGCTC CTCTACACT GACCATCAAA
 2701 CTGCGCTACA AGCCTGATAT CTATCGTGTC AACCCTCACA ATATTGTGAC
 2751 TGTCGTCTCA AACCAAGAGA GCACCTTCGAT CTCAGGAGCA AATCTACGCC
 2801 GCCACGGTTT GTTTGTACAA ATCCATGATG TAGTAGATCT CACCGAGGAC
 2851 ACTCAGGCCT TTCTAAACTA TACCTTTGAC GGGAAAAATG GATTTACAAA
 25 2901 CCACCGAGTG TCTACAGGAC TAAAATCCAC ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 18A. The recombinant protein was used to immunise mice, whose sera were used in an immunoblot analysis blot (Figure 18B) and for FACS analysis (Figure 18C). A his-tagged protein was also expressed.

The cp6737 protein was also identified in the 2D-PAGE experiment (Cpn0454) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6737 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 19

The following *C.pneumoniae* protein (PID 4377090) was expressed <SEQ ID 37; cp7090>:

1 **MNIHSLWKLC TLLALLALPA** CSLSPNYGWE DSCNTCHHTR RKKPSSFGFV
 51 PLYTEEDFNP NPTFGEYDSK BEKQYKSSQV AAFRNITFAT DSYTIKGEEN
 101 LAILTNLVHY MKKNPKATLY IEGHTDERGA ASYNLALGAR RANAIKEHLR
 40 151 KQGISADRLS TISYGKEHPL NSGHNELAWQ QNRRTTEFKIH AR*

A predicted signal peptide is highlighted.

The cp7090 nucleotide sequence <SEQ ID 38> is:

1 ATGAATATAC ATTCCCTATG GAAACTTTGT ACTTTATTGG CTTTACTTGC
 51 ATTGCCAGCA TGTAGCCTTT CCCCTAATTA TGGCTGGGAG GATTCTTGTA
 45 101 ATACATGCCA TCATACAAGA CGAAAAAGC CTTCTTCTTT TGGCTTTGTT
 151 CCTCTCTATA CCGAAGAGGA CTTTAACCCT AATTTTACCT TCGGTGAGTA
 201 TGATTCCAAA GAAGAAAAAC AATACAAGTC AAGCCAAAGT GCAGCATTTT
 251 GTAATATCAC CTTTGTCTACA GACAGCTATA CAATTAAAGG TGAAGAGAAC
 301 CTTGCGATTC TCACGAACCTT GGTTCCTACT ATGAAGAAAA ACCCGAAAGC
 50 351 TACACTGTAC ATTGAAGGGC ATACTGACGA GCGTGGAGCT GCATCCTATA
 401 ACCTTGCTTT AGGAGCACGA CGAGCCAATG CGATTAAAGA GCATCTCCGA
 451 AAGCAGGGAA TCTCTGCAGA TCGTCTATCT ACTATTTCTT ACGGAAAAA

501 ACATCCTTTA AATTCGGGAC ACAACGAACT AGCATGGCAA CAAATCGCC
551 GTACAGAGTT TAAGATTCAT GCACGCTAA

The PSORT algorithm predicts an outer membrane location (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 19A.

- 5 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 19B) and for FACS analysis.

These experiments show that cp7090 is useful immunogen. These properties are not evident from the sequence alone.

Example 20

- 10 The following *C.pneumoniae* protein (PID 4377091) was expressed <SEQ ID 39; cp7091>:

1 MLRQLCFQVF FFCFASLVYA EELEVVVRSE HITLPIEVSC QTDTKDPKIQ
51 KYLSSLTEIF CKDIALGDCL QPTAASKESS SPLAISLRLH VPQLSVVLLQ
101 SSKTPQTLCS FTISQNLSDV RQKIHHAAADT VHYALTGIPG ISAGKIVFAL
151 SSLGKDQKLK QGELWTPDYD GKNLAPLTTE CSLSITPKWV GVGSNFPYLY
15 201 VSYKYGVPKI FLGSLNTEB KKVLPLKGNQ LMPTFSRPRK LLAFVADTYG
251 NPDLFIQPFSLTSGPMGRPR RLLNENFGTQ GNPSFNPEGS QLVFISNKDG
301 RPRLYIMSLD PEPQAPRLLT KKYRNSSCPA WSPDGKKIAF CSVIKGVQRQI
351 CIYDLSSGED YQLTTSPTNK ESPSWAIDSR HLVFSAGNAE ESELYLISLV
401 TKKTNKIAIG VGEKRFPSWG AFPQQPIKRT L*

- 20 A predicted signal peptide is highlighted.

The cp7091 nucleotide sequence <SEQ ID 40> is:

1 ATGTTACGGC AACTATGCTT CCAAGTTTTT TTCTTTTGCT TCGCATCGCT
51 AGTCTATGCT GAAGAATTAG AAGTTGTGTG CCGTTCCGAA CATATCAGCG
101 TCCCTATTGA GGTCTCTTGC CAGACCGATA CGAAAGATCC AAAAATACAG
25 151 AAATACCTCA GCTCGCTAAC GGAGATATTT TGCAAGGACA TTGCCCTAGG
201 AGATTGTCTA CAACCCACAG CGGCTTCTAA AGAATCGTCA TCTCCTTTAG
251 CAATATCTTT ACGGTTGCAT GTACCTCAGC TATCTGTAGT GCTTTTACAG
301 TCTTCAAAA CTCCTCAAAC CTATATGTTCT TTACTATTT CTCAAAATCT
351 TTCTGTAGAT CGTCAAAAA TCCATCACGC TGCTGATACA GTTCATTACG
30 401 CCTCACAGG GATTCCTGGA ATCAGTGCTG GGAAAATTGT TTTTGCTCTA
451 AGTTCTTTAG GAAAAGATCA AAAGCTCAAG CAAGGAGAAT TATGGACTAC
501 AGATTACGAT GGGAAAAACC TCGCCCCCTT AACCACAGAA TGTTCGCTCT
551 CTATAACTCC AAAATGGGTG GGTGTGGGAT CAAATTTTCC CTATCTCTAT
601 GTTTCGTATA AGTATGGTGT GCCTAAAATT TTCTTTGGTT CCTAGAGAA
35 651 CACTGAAGGT AAAAAAGTCC TTCCGTTAAA AGGCAACCAA CTCATGCCTA
701 CGTTTTCTCC AAGAAAAAAG CTTTTAGCTT TCGTTGCTGA TACGTATGGA
751 AATCCTGATT TATTTATTCA ACCGTTCTCA CTAACCTCAG GACCTATGGG
801 TCGCCACGCT CGCCTCCTTA ATGAGAATTT CGGGACTCAA GGAATCCCT
851 CCTTCAACCC TGAAGGATCC CAGCTTGCTT TTATATCGAA CAAAGACGGC
40 901 CGTCCGCGTC TTTATATTAT GTCCCTCGAT CCTGAACCCC AAGCACCTCG
951 CTGCTGACA AAAAAATACA GAAATAGCAG TTGCCCTGCA TGGTCTCCAG
1001 ATGGTAAAAA AATAGCCTTC TGCTCTGTAA TTAAAGGGGT GCGACAAAT
1051 TGTATTACG ATCTCTCCTC TGGAGAGGAT TACCAACTCA CTACGTCTCC
1101 CACAAATAAA GAGAGTCCTT CTTGGGCTAT AGACAGCCGT CATCTTGTCT
45 1151 TTAGTGCGGG GAATGCTGAA GAATCAGAGT TATATTTAAT CAGTCTAGTC
1201 ACCAAAAAAA CTAACAAAAT TGCTATAGGA GTAGGAGAAA AACGGTTCCC
1251 CTCCTGGGGT GCTTTCCTTC AGCAACCGAT AAAGAGAACA CTATGA

The PSORT algorithm predicts an inner membrane location (0.109).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 20A.

- 50 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 20B) and for FACS analysis.

These experiments show that cp7091 is a useful immunogen. These properties are not evident from the sequence alone.

Example 21

The following *C.pneumoniae* protein (PID 4376260) was expressed <SEQ ID 41; cp6260>:

```

5      1  MRFSLCGFPL VFSFTLLSVF DTSLSATTIS LTPEDSFHGD SQNAERSYNV
      51  QAGDVYSLTG DVSISNVDNS ALNKACFNVT SGSVTFAGNH HGLYFNNISS
     101  GTTKEGAVLC QDQPQATARF SGFSTLSFIQ SPGDIKEQGC LYSKNALMLL
     151  NNYVVRFEQN QSKTKGGAIS GANVTIVGNY DSVSFYQNA TFGGAHSSG
     201  PLQIAVNOAE IRFAQNTAKN GSGGALYSDG DIDIDQNAVY LFRENEALTT
     251  AIGKGGAVCC LPTSGSSTPV PIVTFSDNKQ LVFERNHSIM GGGAIYARKL
     301  SISSGGPTLF INNISYANSQ NLGGAIAIDT GGEISLSAEK GTITTFQGNRT
     351  SLPFLNGIHL LQNAKFLKLQ ARNGYSIEFY DPITSEADGS TQLNNGDPK
     401  NKEYTGTILF SGEKSLANDP RDFKSTIPQN VNLSAGYLVI KEGAEVTVSK
     451  FTQSPGSHLV LDLGTKLIAS KEDIAITGLA IDIDSLSSSS TAAVIKANTA
     501  NKQISVTDSE ELISPTGNAY EDLRMRNSQT FLLSLEPGA GGSVTVTAGD
     551  FLPVSPHYGF QGNWKLAWTG TGNKVGEFFW DKINYKPRPE KEGNLVFNIL
     601  WGNADVDRSL MQVQETHASS LQTDRLWID GIGNFFHVSA SEDNIRYRHN
     651  SGGYVLSVNN EITPKHYTSM AFSQLFSRDK DYAVSNNEYR MYLGSYLYQY
     701  TTS LGNIFRY ASRNPVNVNG ILSRRFLQNP LMIFHFLCAY GHATNDMKTG
     751  YANFPMVKNS WRNNCWAIEC GGSMPLLVFE NGRLFQGAIP FMKLQLVYAY
     801  QGDFKETAD GRRFSNGSLT SISVPLGIRF EKLALSQDVL YDFSFSYIPD
     851  IFRKDPSCEA ALVISGDSWL VPAAHVSRHA FVSGTGRYH FNDYTELLCR
     901  GSIECRPHAR NYNINCGSKF RF*

```

A predicted signal peptide is highlighted.

25 The cp6260 nucleotide sequence <SEQ ID 42> is:

```

      1  ATGCGATTTT CGCTCTGCGG ATTTCCTCTA GTTTTTTCTT TTACATTGCT
      51  CTCAGTCTTC GACACTTCTT TGAGTGCTAC TACGATTTCCT TTAACCCAG
     101  AAGATAGTTT TCATGGAGAT AGTCAGAATG CAGAACGTTT TTATAATGTT
     151  CAAGCTGGGG ATGCTCTATAG CCTTACTGGT GATGCTCTCA TATCTAACGT
     201  CGATAAATCT GCATTAAATA AAGCCTGCTT CAATGTGACC TCAGGAAGTG
     251  TGACGTTCGC AGGAAATCAT CATGGGTAT ATTTTAATAA TATTTCCTCA
     301  GGAACACAA AGGAAGGGGC TGTACTTTGT TGCCAAGATC CTCAGCAAC
     351  GGCACGTTT TCTGGGTCT CCACGCTCTC TTTTATTCAG AGCCCCGGAG
     401  ATATTAAAGA ACAGGGATGT CTCTATTCAA AAAATGCACT TATGCTCTTA
     451  AACAAATTATG TAGTGCCTTT TGAACAAAAC CAAAGTAAGA CTAAGGCGG
     501  AGCTATTAGT GGGGCGAATG TTACTATAGT AGGCAACTAC GATTCCTCT
     551  CTTTCTATCA GAATGCAGCC ACTTTTGGAG GTGCTATCCA TTCTTCAGGT
     601  CCCCTACAGA TTGCAGTAAA TCAGGCAGAG ATAAGATTG CACAAAATAC
     651  TGCCAAGAAT GGTCTGAGG GGGCTTTGTA CTCCGATGGT GATATTGATA
     701  TTGATCAGAA TGCTTATGTT CTATTTCGAG AAAATGAGGC ATTGACTACT
     751  GCTATAGGTA AGGGAGGGGC TGTCTGTTGT CTTCCACATT CAGGAAGTAG
     801  TACTCCAGTT CCTATTGTGA CTTTCTCTGA CAATAAACAG TTAGTCTTTG
     851  AAAGAAACCA TTCCATAATG GGTGGCGGAG CCATTTATGC TAGGAACTT
     901  AGCATCTCTT CAGGAGGTCC TACTCTATTT ATCAATAATA TATCATATGC
     951  AAATTCGCAA AATTTAGGTG GAGCTATTGC CATTGATACT GGAGGGGAGA
    1001  TCAGTTTATC AGCAGAGAAA GGAACAATTA CATTCCAAGG AAACCGGACG
    1051  AGCTTACCGT TTTTGAATGG CATCCATCTT TTACAAAATG CTAATTCCT
    1101  GAAATTACAG GCGAGAAATG GATACTCTAT AGAATTTTAT GATCCTATTA
    1151  CTTCTGAAGC AGATGGGTCT ACCCAATTGA ATATCAACGG AGATCCTAAA
    1201  AATAAAGAGT ACACAGGGAC CATACTCTTT TCTGGAGAAA AGAGTCTAGC
    1251  AAACGATCCT AGGATTTTAA AATCTACAAT CCCTCAGAAC GTCAACCTGT
    1301  CTGCAGGATA CTTAGTTATT AAAGAGGGGG CCGAAGTCAC AGTTTCAAAA
    1351  TTCACGCAGT CTCCAGGATC GCATTTAGTT TTAGATTTAG GAACCAAACT
    1401  GATAGCCTCT AAGGAAGACA TTGCCATCAC AGGCCTCGCG ATAGATATAG
    1451  ATAGCTTAAG CTCATCCTCA ACAGCAGCTG TTATTAAAGC AAACACCGCA
    1501  AATAAACAGA TATCCGTGAC GGACTCTATA GAACTTATCT CGCCTACTGG
    1551  CAATGCCTAT GAAGATCTCA GAATGAGAAA TTCACAGACG TTCCCTCTGC
    1601  TCTCTTTAGA GCCTGGAGCC GGGGGTAGTG TGAAGTAAAC TGCTGGAGAT
    1651  TTCCTACCGG TAAGTCCCCA TTATGGTTTT CAAGGCAATT GGAAATTAGC
    1701  TTGGACAGGA ACTGGAACAA AAGTTGGAGA ATTCTTCTGG GATAAAATAA

```

1751 ATTATAAGCC TAGACCTGAA AAAGAAGGAA ATTTAGTTCC TAATATCTTG
 1801 TGGGGGAATG CTGTAGATGT CAGATCCTTA ATGCAGGTTT AAGAGACCCA
 1851 TGCATCGAGC TTACAGACAG ATCGAGGGCT GTGGATCGAT GGAATTGGGA
 1901 ATTTCTTCCA TGTATCTGCC TCCGAAGACA ATATAAGGTA CCGTCATAAC
 1951 AGCGGTGGAT ATGTTCTATC TGTAAATAAT GAGATCACAC CTAAGCACTA
 2001 TACTTCGATG GCATTTTCCC AACTCTTTAG TAGAGACAAG GACTATGCGG
 2051 TTCCAACAA CGAATACAGA ATGTATTTAG GATCGTATCT CTATCAATAT
 2101 ACAACCTCCC TAGGGAATAT TTTCCGTTAT GCTTCGCGTA ACCCTAATGT
 2151 AAACGTCGGG ATTCTCTCAA GAAGGTTTCT TCAAAATCCT CTATGATTT
 2201 TTCATTTTTF GTGTGCTTAT GGTATGCCA CCAATGATAT GAAAACAGAC
 2251 TACGCAAAAT TCCCTATGGT GAAAAACAGC TGGAGAAACA ATGTGTGGGC
 2301 TATAGAGTGC GGAGGGAGCA TGCCTCTATT GGTATTTGAG AACGGAAGAC
 2351 TTTTCCAAGG TGCCATCCCA TTTATGAAAC TACAATTAGT TTATGCTTAT
 2401 CAGGGAGATT TCAAAGAGAC GACTGCAGAT GGCCGTAGAT TTAGTAATGG
 2451 GAGTTTAAAC TCGATTCTCG TACCTCTAGG CATAAGCTTT GAGAAGCTGG
 2501 CACTTTCTCA GGTATGACTC TATGACTTTA GTTCTCTCTA TATTCTGAT
 2551 ATTTTCCGTA AGGATCCCTC ATGTGAAGCT GCTCTGGTGA TTAGCGGAGA
 2601 CTCCTGGCTT GTTCCGGCAG CACACGTATC AAGACATGCT TTGTAGGGA
 2651 GTGGAACGGG TCGGTATCAC TTTAACGACT ATACTGAGCT CTTATGTCGA
 2701 GGAAGTATAG AATGCCGCC CCATGCTAGG AATTATAATA TAAACTGTGG
 2751 AAGCAAATTT CGTTTTTAG

The PSORT algorithm predicts an outer membrane location (0.921).

The protein was expressed in *E. coli* and purified both as a his-tag and GST-fusion product. The GST-fusion is shown in Figure 21A. This recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 21B) and for FACS analysis (Figure 21C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6260 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 22

The following *C. pneumoniae* protein (PID 4376456) was expressed <SEQ ID 43; cp6456>:

1 MSSPVNTPS APNIPAPPT TPGIPTTKPR SSFIEKVIIV AKYILFAIAA
 51 TSGALGTILG LSGALTFGIG IALLVIFVVS MVLLGLILKD SISGGEERRL
 101 REEVSRTSE NQRLTVITTT LETEVKDLKA AKDQLTLEIE AFRNENGNLK
 151 TTAEDLEEQV SKLSEQLEAL ERINQLIQAN AGDAQEISSE LKKLISGWDS
 201 KVVEQINTSI QALKVLLGQE WVQEAQTHVK AMQEIQALQ AEILGMHNQS
 251 TALQKSVENL LVQDQALTRV VGELLESENK LSQACSALRQ EIEKLAQHET
 301 SLQQRIDAML AQEQNLAEQV TALEKMKQEA QKAESEFIAC VRDRTFGRRE
 351 TPPPPTPVVE GDESQEEDG GTPPVSQPS PVDRTGDDGQ *

The cp6456 nucleotide sequence <SEQ ID 44> is:

1 ATGTCATCTC CTGTAAATAA CACACCCTCA GCACCAACA TTCCAATACC
 51 AGCGCCACG ACTCCAGGTA TTCCTACAAC AAAACCTCGT TCTAGTTTCA
 101 TTGAAAAGGT TATCATTTGTA GCTAAGTACA TACTATTTGC AATTGCAGCC
 151 ACATCAGGAG CACTCGGAAC AATTCTAGGT CTATCTGGAG CGCTAACCCC
 201 AGGAATAGGT ATTGCCCTTC TTGTTATCTT CTTTGTTCCT ATGGTGCTTT
 251 TAGGTTTAAT CCTTAAAGAT TCTATAAGTG GAGGAGAAGA ACGCAGGCTC
 301 AGAGAAGAGG TCTCTCGATT TACAAGTGAG AATCAACGGT TGACAGTCAT
 351 AACCACAACA CTTGAGACTG AAGTAAAGGA TTTAAAGCA GCTAAAGATC
 401 AACTTACACT TGAATCGAA GCATTTAGAA ATGAAAACGG TAATTTAAAA
 451 ACAACTGCTG AGGACTTAGA AGAGCAGGTT TCTAACTTA CGGAACAATT
 501 AGAAGCACTA GAGCGAATTA ATCAACTTAT CCAAGCAAAC GCTGGAGATG
 551 CTCAAGAAAT TTCGTCTGAA CTAAAGAAAT TAATAAGCGG TTGGGATTCC
 601 AAAGTTGTTG AACAGATAA TACTTCTATT CAAGCATTTA AAGTGTATT
 651 GGGTCAAGAG TGGGTGCAAG AGGCTCAAAC ACACGTTAAA GCAATGCAAG
 701 AGCAAATTCA AGCATTGCAA GCTGAAATTC TAGGAATGCA CAATCAATCT

5
 751 ACAGCATTCG AAAAGTCAGT TGAGAATCTA TTAGTACAAG ATCAAGCTCT
 801 AACAAAGAGTA GTAGGTGAGT TGTTAGAGTC TGAGAACAAAG CTAAGCCAAG
 851 CTTGTTCTGC GCTACGTCAA GAAATAGAAA AGTTGGCCCA ACATGAAACA
 901 TCTTTGCAAC AACGTATTGA TCGGATGCTA GCCCAAGAGC AAAATTGGC
 951 AGAGCAGGTC ACAGCCCTTG AAAAAATGAA ACAAGAAGCT CAGAAGGCTG
 1001 AGTCCGAGTT CATTGCTTGT GTACGTGATC GAACCTTCGG ACGTCGTGAA
 1051 ACACCTCCAC CAACAACACC TGTAAGTGAA GGTGATGAAA GTCAAGAAGA
 1101 AGACGAAGGA GGTACTCCCC CAGTATCACA ACCATCTTCA CCCGTAGATA
 1151 GAGCAACAGG AGATGGTCAG TAA

10 The PSORT algorithm predicts inner membrane (0.127).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 22A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 22B) and for FACS analysis (Figure 22C). A his-tag protein was also expressed.

15 These experiments show that cp6456 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 23

The following *C.pneumoniae* protein (PID 4376729) was expressed <SEQ ID 45; cp6729>:

1 MKIPLHKLLI SSTLVTPILL SIATYGADAS LSPTDSFDGA GGSTFFPKST
 51 ADANGTNVVL SGNVYINDAG KGTALTGCCF TETTGDLTFT GKGYSFSPNT
 101 VDAGSNAGAA ASTTADKALT FTGFSNLSFI AAPGTTVASG KSTLSSAGAL
 151 NLFDNGTILF SQNVSNNEAN NGGAIITTKL SLSGNTSSIT FTSNSAKKLG
 201 GAIYSSAAAS ISGNTGQLVF MNNKGETGGG ALGFEASSSI TQNSSLFFSG
 251 NTATDAAGKG GAIYCEKTGE TPTLTISGNK SLTFAENSSV TQGGAIKAHG
 301 LDLSAAGPTL FSNNRGNTA AGKGAIAIA DSGSLSLSAN QGDITFLGNT
 351 LTSTSAPTST RNAIYLSSA KITNLRAAQG QSIYFYDPIA SNTTGASDVL
 401 TINQPDNSNP LDYSGTIVFS GEKLSADEAK AADNFTSILK QPLALASGTL
 451 ALKGNVELDV NGFTQTEGST LLMQPGTKLK ADTEAISLTK LVVDLSALEG
 501 NKSVSIEITAG ANKTITLTSP LVFQDSSGNF YESHTINQAF TQPLVVFTAA
 551 TAASDIYIDA LLTSPVQTPE PHYGYQGHWE ATWADTSTAK SGTMTVWTTG
 601 YNPNPERRAS VVPDSLWASF TDIRTLQQIM TSQANSIYQQ RGLWASGTAN
 651 FFHKDKSGTN QAFRHKSIVGY IVGGSIEDFS ENIFSVAFCQ LFGKDKDLFI
 701 VENTSHNYLA SLYLQHRAFL GGLPMPSPGS ITDMLKDIPL ILNAQLSVSY
 751 TKNDMDTRYT SYPEAQGSWT NNSGALELGG SLALYLPKEA PFFQGYFPFL
 801 KFQAVYSRQQ NFKESGAEAR AFDGDLVNC SIPVGIKLEK ISEDEKNFFE
 851 ISLAYIGDVY RKNPRSRISL MVSGASWTSI CKNLARQAFL ASAGSHLTLS
 901 PHVELSGEAA YELRGSATY NVDCGLRYSF *

A predicted signal peptide is highlighted.

The cp6729 nucleotide sequence <SEQ ID 46> is:

40
 1 ATGAAAATAC CCTTGACAA ACTCCTGATC TCTTCGACTC TTGTCACCTC
 51 CATTCTATTG AGCATTGCAA CTTACGGAGC AGATGCTTCT TTATCCCCCTA
 101 CAGATAGCTT TGATGGAGCG GCGGCTCTA CATTACTACC AAAATCTACA
 151 GCAGATGCCA ATGGAACGAA CTATGTCTTA TCAGGAAATG TCTATATAAA
 201 CGATGCTGGG AAAGGCACAG CATTAACAGG CTGCTGCTTT ACAGAAACTA
 251 CGGGTGATCT GACATTTACT GGAAAGGGAT ACTCATTTTC ATTCAACACG
 45
 301 GTAGATGCGG GTTCGAATGC AGGAGCTGCG GCAAGCACAA CTGCTGATAA
 351 AGCCCTAACA TTCACAGGAT TTTCTAACCT TTCCTTCATT GCAGCTCCTG
 401 GAACACAGT TGCTTCAGGA AAAAGTACTT TAAGTTCCTG AGGAGCCTTA
 451 AATCTTACCG ATAATGGAAC GATTCTCTTT AGCCAAAACG TCTCCAATGA
 50
 501 AGCTAATAAC AATGGCGGAG CGATCACCAC AAAAATCTTT TCTATTCTG
 551 GGAATACCTC TTCTATAACC TTCACTAGTA ATAGCGCAA AAAATTAGGT
 601 GGAGCGATCT ATAGCTCTGC GGCTGCAAGT ATTTTCAGGAA ACACCGGCCA
 651 GTTAGTCTTT ATGAATAATA AAGGAGAAAC TGGGGGTGGG GCTCTGGGCT
 701 TTGAAGCCAG CTCCTCGATT ACTCAAAATA GCTCCCTTTT CTCTCTGGA
 751 AACACTGCAA CAGATCTGCG AGGCAAGGGC GGGGCCATTT ATTGTGAAAA
 801 AACAGGAGAG ACTCCTACTC TTACTATCTC TGGAAATAAA AGTCTGACCT
 851 TCGCCGAGAA CTCTTCAGTA ACTCAAGGCG GAGCAATCTG TGCCCATGCT

5
 10
 15
 20
 25
 30
 35

```

901 CTAGATCTTT CCGCTGCTGG CCCTACCTTA TTTTCAAATA ATAGATGCGG
951 GAACACAGCT GCAGGCAAGG GCGGCGCTAT TGCAATTGCC GACTCTGGAT
1001 CTTTAAGTCT CTCTGCAAAT CAAGGAGACA TCACGTTCCCT TGGCAACACT
1051 CTAACCTCAA CCTCCGCGCC AACATCGACA CGGAATGCTA TCTACCTGGG
1101 ATCGTCAGCA AAAATTACGA ACTTAAGGGC AGCCCAAGGC CAATCTATCT
1151 ATTTCTATGA TCCGATTGCA TCTAACACCA CAGGAGCTTC AGACGTCTCTG
1201 ACCATCAACC AACCGGATAG CAACTCGCCT TTAGATTATT CAGGAACGAT
1251 TGTATTTTCT GGGGAAAAGC TCTCTGCAGA TGAAGCGAAA GCTGCTGATA
1301 ACTTCACATC TATATTAAAG CAACCATTGG CTCTAGCCTC TGAACCTTA
1351 GCACTCAAAG GAAATGTCGA GTTAGATGTC AATGGTTTCA CACAGACTGA
1401 AGGCTCTACA CTCCTCATGC AACCAGGAAC AAAGCTCAAA GCAGATACTG
1451 AAGCTATCAG TCTTACCAA CTGTCTGTTG ATCTTTCTGC CTTAGAGGGA
1501 AATAAGAGTG TGTCCATTGA AACAGCAGGA GCCAACAAAA CTATAACTCT
1551 AACCTCTCCT CTTGTTTTC AAGATAGTAG CGCAATTTT TATGAAGCC
1601 ATACGATAAA CCAAGCCTTC ACGCAGCCTT TGGTGGTATT CACTGCTGCT
1651 ACTGCTGCTA GCGATATTTA TATCGATGCG CTCTCTACTT CTCCAGTACA
1701 AACTCCAGAA CCTCATTACG GGTATCAGGG ACATTGGGAA GCCACTTGGG
1751 CAGACACATC AACTGCAAAA TCAGGAACCTA TGACTTGGGT AACTACGGGC
1801 TACAACCCTA ATCCTGAGCG TAGAGCTTCC GTAGTTCCTG ATTCATTATG
1851 GGCATCCTTT ACTGACATTC GCACTCTACA GCAGATCATG ACATCTCAAG
1901 CGAATAGTAT CTATCAGCAA CGAGGACTCT GGGCATCAGG AACTGCGAAT
1951 TTCTTCCATA AGGATAAATC AGGAACCTAAC CAAGCATTCC GACATAAAAG
2001 CTACGGCTAT ATTGTTGGAG GAAGTGCTGA AGATTTTCTT GAAAATATCT
2051 TCAGTGTAGC TTTCTGCCAG CTCTTCGGTA AAGATAAAGA CCTGTTTATA
2101 GTTGAAAATA CCTCTCATAA CTATTTAGCG TCGCTATACC TGCAACATCG
2151 AGCATTCCTA GGAGGACTTC CCATGCCCTC ATTTGGAAGT ATCACCAGACA
2201 TGCTGAAAGA TATTCCTCTC ATTTTGAATG CCCAGCTAAG CTACAGCTAC
2251 ACTAAAAATG ATATGGATAC TCGCTATACT TCCTATCCTG AAGCTCAAGG
2301 CTCTTGGACC AATAACTCTG GGGCTCTAGA GCTCGGAGGA TCTCTGGCTC
2351 TATATCTCCC TAAAGAAGCA CCGTTCCTCC AGGGATATTT CCCCTTCTTA
2401 AAGTTCACAG CAGTCTACAG CCGCCAACAA AACTTTAAAG AGAGTGGCGC
2451 TGAAGCCCGT GCTTTTGATG ATGGAGACCT AGTGAAGTGC TCTATCCCTG
2501 TCGGCATTCG GTTAGAAAAA ATCTCCGAAG ATGAAAAAAA TAATTTGAG
2551 ATTTCTCTAG CCTACATTGG TGATGTGTAT CGTAAAAATC CCCGTTGCGC
2601 TACTTCTCTA ATGGTCAGTG GAGCCTCTTG GACTTCGCTA TGTAAAAAC
2651 TCGCACGACA AGCCTTCTTA GCAAGTGCTG GAAGCCATCT GACTCTCTCC
2701 CCTCATGTAG AACTCTCTGG GGAAGCTGCT TATGAGCTTC GTGGCTCAGC
2751 ACACATCTAC AATGTAGATT GTGGGCTAAG ATACTATTTC TAG
  
```

The PSORT algorithm predicts outer membrane (0.927).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 23A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 23B) and for FACS analysis (Figure 23C). A his-tag protein was also expressed.

The cp6729 protein was also identified in the 2D-PAGE experiment (Cpn0446) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

45 These experiments show that cp6729 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 24

The following *C.pneumoniae* protein (PID 4376849) was expressed <SEQ ID 47; cp6849>:

50
 55

```

1  MSKLIRRVVT VLALTSMA SC FASGGIEAAV AESLITKIVA SAETKPAPVP
51  MTAKKVLRLV RNKQPV EQKS RGAFCDKIFY PCEEGRCQPV EAQQESCYGR
101 LYSVKVND D NVEICQSVPE YATVGSPIPI EILAIGKKDC VDVVITQQLP
151 CEAEFVSSDP ETTPTS DSKL VWKIDRLGAG DKCKITVWVK FLKEGCCFTA
201 ATVCACPELR SYTKCGQPAI CIKQEGPDCA CLRCPCVYKI EVVNTGSAIA
251 RNVTVDPNPV DGYSHASGQR VLSFNLGDMR PGDKKVFTVE FCPQRRGQIT
301 NVATVTYCGG HKCSANVT TV VNEPCVQVNI SGADWSYVCK PVEYSISVSN
351 PGDLVLHDVV IQDTLPSGVT VLEAPGGEIC CNKVWRIKE MCPGETLQFK
  
```

401 LVVKAQVPGR FTNQVAVTSE SNCGTCTSCA ETTTHWKGLA ATHMCVLDTN
 451 DPICVGENTV YRICVTNRGS AEDTNVSLIL KFSKELQPIA SSGPTKGTIS
 501 GNTVVPDALP KLGSKESEVF SVTLKGIAPG DARGEAILSS DTLTSPVSDT
 551 ENTHVY*

5 A predicted signal peptide is highlighted.

The cp6849 nucleotide sequence <SEQ ID 48> is:

1 ATGTCCAAAC TCATCAGACG AGTAGTTACG GTCCTTGCGC TAACGAGTAT
 51 GCGGAGTTGC TTTGCCAGCG GGGGTATAGA GGCCGCTGTA GCAGAGTCTC
 101 TGATTACTAA GATCGTCGCT AGTGCAGAAA CAAAGCCAGC ACCGTTCCT
 151 ATGACAGCGA AGAAGGTTAG ACTTGTCGGT AGAAATAAAC AACCAGTTGA
 201 ACAAAAAGC CGTGGTGCTT TTTGTGATAA AGAATTTTAT CCGTGTGAAG
 251 AGGACGATG TCAACCTGTA GAGGCTCAGC AAGAGTCTTG CTACGGAAGA
 301 TTGTATCTCG TAAAAGTAAA CGATGATTGC AACGTAGAAA TTTGCCAGTC
 351 CGTCCAGAA TACGCTACTG TAGGATCTCC TTACCCTATT GAAATCCTTG
 15 401 CTATAGGCAA AAAAGATTGT GTTGATGTTG TGATTACACA ACAGCTACCT
 451 TCGGAAGCTG AATTCTGTAAG CAGTGATCCA GAAACAATC CTACAAGTGA
 501 TGGGAAATTA GTCTGGAAAA TCGATCGCCT GGGTGCAGGA GATAAATGCA
 551 AAACTACTGT ATGGGTAAAA CCTCTTAAAG AAGGTGCTG CTTACAGCT
 601 GCTACTGTAT GTGCTTGCCC AGAGCTCCGT TCTTATACTA AATGCGGTCA
 20 651 ACCAGCCATT TGTATTAAGC AAGAAGGACC TGACTGTGCT TGCCTAAGAT
 701 GCCCTGTATG CTACAAA TCGAAGTAGTGA ACACAGGATC TGCTATTGCG
 751 CGTAACGTAA CTGTAGATAA TCCTGTTCCC GATGGCTATT CTCATGCTATC
 801 TGGTCAAAGA GTTCTCTCTT TTAACCTAGG AGACATGAGA CCTGGCGATA
 851 AAAAGGTATT TACAGTTGAG TTCTGCCCTC AAAGAAGAGG TCAAAATCACT
 25 901 AACGTTGCTA CTGTAACCTA CTGCGGTGGA CACAAATGTT CTGCAATGT
 951 AACTACAGTT GTTAATGAGC CTTGTGTACA AGTAAATATC TCTGGTGCTG
 1001 ATTGGTCTTA CGTATGTAAA CCTGTGGAGT ACTCTATCTC AGTATCGAAT
 1051 CCTGGAGACT TGGTTCTTCA TGATGTCGTG ATCCAAGATA CACTCCCTTC
 1101 TGGTGTTACA GTACTCGAAG CTCCTGGTGG AGAGATCTGC TGTATAAAG
 30 1151 TTGTTTGGCG TATTAAGAA ATGTGCCCAG GAGAAACCTC CCAGTTTAAA
 1201 CTTGTAGTGA AAGCTCAAGT TCCTGGAAGA TTCACAAATC AAGTTGCAGT
 1251 AACTAGTGAG TCTAACTGCG GAACATGTAC ATCTTGCGCA GAAACAACAA
 1301 CACATTGGAA AGGTCTTGCA GCTACCCATA TGTGCGTATT AGACACAAAT
 1351 GATCCTATCT GTGTAGGAGA AAATACTGTC TATCGTATCT GTGTAACATA
 35 1401 CCGTGGTTCT GCTGAAGATA CTAACGTATC TTTAATCTTG AAGTTCTCAA
 1451 AAGAACTTCA GCCAATAGCT TCTTCAGGTC CAACTAAAGG AACGATTTC
 1501 GGTAAATACCG TTGTTTTCGA CGCTTTACCT AAACCTCGGT CTAAGGAATC
 1551 TGTAAGATT TCTGTTACCT TGAAAGGTAT TGCTCCCGGA GATGCTCGCG
 1601 GCGAAGCTAT TCTTCTTCT GATACACTGA CTTACACAGT ATCAGACACA
 40 1651 GAAATACCC ACGTGATTA A

The PSORT algorithm predicts periplasmic space (0.93).

The protein was expressed in *E. coli* and purified as a GST-fusion product, as shown in Figure 24A, and also as a his-tag protein. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 24B) and for FACS analysis (Figure 24C).

45 The cp6849 protein was also identified in the 2D-PAGE experiment (Cpn0557).

These experiments show that cp6849 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 25

The following *C. pneumoniae* protein (PID 4376273) was expressed <SEQ ID 49; cp6273>:

50 1 MGLFHLTLFG LLLCSLPISL VAKFPESVGH KILYISTQST QOALATYLEA
 51 LDAYGDHDFV VLRKIGEDYL QSIHSSDPQ TRKSTIIGAG LAGSSEALDV
 101 LSQAMETADP LQQLLVLSAV SGHLGKTSDD LLFKALASPY PVIRLEAAYR
 151 LANLKNKTKVI DHLHSFIHKL PEEIQCLSAA IFLRLETEES DAYIRDLLAA
 201 KKSARSATA LQIGEQQKR FLPTLRNLLT SASPDQDEAI LYALGKLDKG

5 251 QSYNKKQL QKPDVDVTLA AAQALIALGK EEDALPVIKK QALEERPRAL
 301 YALRHLPSFI GPIALPIFL KTKNSEAKLN VALALLELGC DTPKLLLEYIT
 351 ERLVQPHYNE TLALSFSKGR TLQNWKRVI IVPQDPQERE RLLSTTRGLE
 401 EQILTFLFRL PKEAYLPCYI KLLASQKTQL ATTAISFLSH TSHQEAIDLL
 451 FQAAKLPGEP IIRAYADLAI YNLTKDPEKK RSLHDYAKKL IQETLLFPVDT
 501 ENQRPHPSMP YLRYQVTPES RTKLMLDILE TLATSKSSED IRLLIQLMTE
 551 GDAKNFPVLA GLLIKIVE*

A predicted signal peptide is highlighted.

The cp6273 nucleotide sequence <SEQ ID 50> is:

10 1 ATGGGACTAT TCCATCTAAC TCTCTTTGGA CTTTTATTGT GTAGTCTTCC
 51 CATTTCTCTT GTTGCTAAAT TCCCTGAGTC TGTAGGTCAT AAGATCCTTT
 101 ATATAAGTAC GCAATCTACA CAGCAGGCGT TAGCAACATA TCTGGAAGCT
 151 CTAGATGCC TACGGTATCA TGAATCTCTT GTTTTAAGAA AAATCGGAGA
 201 AGACTATCTC AAGCAAAGCA TCCACTCCTC AGATCCGCAA ACTGAAAAA
 15 251 GCACCATCAT TGGAGCAGGC CTGGCGGGAT CTTCAGAAGC CTGGACGCTG
 301 CTCTCCCAAG CATATGGAAC TGCAGACCCC CTGCAGCAGC TACTGGTTTT
 351 ATCGGCAGTC TCAGGACATC TTGGGAAAAA TTCTGACGAC TTACTGTTTA
 401 AAGCTTTAGC ATCTCCCTAT CCTGTCATCC GCTTAGAAGC CGCCTATAGA
 451 CTGCTAATT TGAAGAACAC TAAAGTCATT GATCATCTAC ATTCTTTTCAT
 20 501 TCATAAGCTT CCCGAAGAAA TCCAATGCCT ATCTGCGGCA ATATTCTTAC
 551 GCTTGGAGAC TGAAGAATCT GATGCTTATA TTGGGGATCT CTTAGCTGCC
 601 AAGAAAAGCG CGATTGGAG TGCACAGCT TTGCAGATCG GAGAAATACCA
 651 ACAAAAACGC TTTCTTCCGA CACTTAGGAA TTTGCTAACG AGTGCCTCTC
 701 CTCAAGATCA AGAAGCTATT CTTTATGCTT TAGGGAAGCT TAAGGATGGT
 25 751 CAGAGCTACT ACAATATAAA AAAGCAATTG CAGAAGCCTG ATGTGGATGT
 801 CACTTTAGCA GCAGCTCAAG CTTTAATTGC TTTGGGGAAA GAAGAGGACG
 851 CTCTTCCCGT GATAAAAAAG CAAGCACTTG AGGAGCGGCC TCAGAGCCTG
 901 TATGCCCTAC GGCATCTACC CTCTGAGATA GGGATTCCGA TTGCCCTGCC
 951 GATATTCTTA AAAACTAAGA ACAGCGAAGC CAAGTTGAAT GTAGCTTTAG
 30 1001 CTCTCTTAGA GTTAGGGTGT GACACCCCTA AACTACTGGA ATACATTACC
 1051 GAAAGGCTTG TCCAACCACA TTATAATGAG ACTCTAGCCT TGAGTTCTCT
 1101 TAAGGGGCGT ACTTTACAAA ATTGGAAGCG GGTGAACATC ATAGTCCCTC
 1151 AAGATCCCCA GGAGAGGGAA AGGTTGCTCT CCACAACCCG AGGTCTTGAA
 1201 GAGCAGATCC TTACGTTTCT CTTCCGCCTA CCTAAAGAAG CTTACCTCCC
 35 1251 CTGTATTAT AAGCTTTTGG CGAGTCAGAA AACTCAGCTT GCCACTACTG
 1301 CGATTTCTTT TTTAAGTCAC ACCTCACATC AGGAAGCCTT AGATCTACTT
 1351 TTCCAAGCTG CGAAGCTTCC TGGAGAACCT ATCATCCGCG CCTATGCAGA
 1401 TCTTGCTATT TATAATCTCA CCAGAGATCC TGAAAAA AAA CGTTCTCTCC
 1451 ATGATTATGC AAAAAAGCTA ATTCAAGAAA CTTGTTTATT TGTGGACACG
 40 1501 GAAAACCAAA GACCCCATCC CAGCATGCCC TATCTACGTT ATCAGGTCAC
 1551 CCCAGAAAGC CGTACGAAGC TCATGTTGGA TATTCTAGAG AACTAGCCA
 1601 CCTCGAAGTC TTCCGAAGAT ATCCGTTTAT TGATACAAC TATGACGGAA
 1651 GGAGATGCAA AAAATTTCCT AGTCCTTGCA GGCTTACTCA TAAAAATTGT
 1701 GGAGTAA

45 The PSORT algorithm predicts a periplasmic location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 25A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 25B) and for FACS analysis (Figure 25C).

50 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6273 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 26

The following *C.pneumoniae* protein (PID 4376735) was expressed <SEQ ID 51; cp6735>:

1 MTILRNFLTC SALFLALPAA AQVVYLHESD GYNGAINNKS LEPKITCYPE
 51 GTSYIFLDDV RISNVKHDQE DAGVFINRSG NLFFMGNRCN FTFHNLMTGEG
 101 FGAAISNRVG DTTLTLSNFS YLAFTSAPLL PQGQGAISL GSVMIENSEE
 151 VTFCGNYSSW SGAAIYTPYL LGSKASRPSV NLSGNRYLVF RDNVSQGYGG
 201 AISTHNLTLT TRGPSCFENN HAYHDVNSNG GAIAIAPGGS ISISVKSDDL
 251 IFKGNTASQD GNTIHNSIHL QSGAQFKNLR AVSESGVYFY DPISHSESHK
 301 ITDLVINAPE GKETYEGTIS FSGLCDDHE VCAENLTSTI LQDVTLAGGT
 351 LSLSDGVTLQ LHSFKQEASS TLTMSPGTTL LCSGDARVQN LHILIEDTDN
 401 FVPVRIRAED KDALVSLEKL KVAFEAYWSV YDFPQFKEAF TIPLLELLGP
 451 SFDSLLLGET TLERTQVTE NDAVRGFWSL SWEEYPPSLD KDRRITPTKK
 501 TVFLTWNPEI TSTP*

A predicted signal peptide is highlighted.

The cp6735 nucleotide sequence <SEQ ID 52> is:

1 ATGACCATAC TTCGAAATTT TCTTACCTGC TCGGCTTTAT TCCTCGCTCT
 51 CCCTGCAGCA GCACAAGTTG TATATCTTCA TGAAAGTGAT GGTATATAACG
 101 GTGCTATCAA TAATAAAGC TTAGAACCTA AAATTACCTG TTATCCAGAA
 151 GGAACCTCTT ACATCTTTCT AGATGACGTG AGGATTTCCA ACCTTAAGCA
 201 TGATCAAGAA GATGCTGGGG TTTTATAAAA TCGATCTGGG AATCTTTTTT
 251 TCATGGGCAA CCGTTGCAAC TTCACCTTTC ACAACCTTAT GACCGAGGGT
 301 TTTGGCGCTG CCATTTTCGAA CCGCGTTGGA GACACCACTC TCACTCTCTC
 351 TAAATTTTCT TACTTAGCGT TCACCTCAGC ACCTCTACTA CCTCAAGGAC
 401 AAGGAGCGAT TTATAGTCTT GGTTCCTGTA TGATCGAAAA TAGTGAGGAA
 451 GTGACTTTCT GTGGGAATA CTCTTCGTGG AGTGGAGCTG CGATTTATAC
 501 TCCCTACCTT TTAGGTTCTA AGGCGAGTCG TCCTTCAGTA AATCTCAGCG
 551 GGAACCGCTA CTTGGTGTTT AGAGACAATG TGAGCCAAGG TTATGGCGCG
 601 GCCATATCTA CCCACAATCT CACACTCAGC ACTCGAGGAC CTTCTGTGTT
 651 TGAAATAAAT CATGCTTATC ATGACGTGAA TAGTAATGGA GGAGCCATTG
 701 CCATTGCTCC TGGAGGATCG ATCTCTATAT CCGTGAAAAG CCGAGATCTC
 751 ATCTTCAAAG GAAATACAGC ATCACAAGAC GGAAATACAA TACACAATC
 801 CATCCATCTG CAATCTGGAG CACAGTTTAA GAACCTACGT GCTGTTTCAG
 851 AATCCGGAGT TTATTTCTAT GATCCTATAA GCCATAGCGA GTCGCATAAA
 901 ATTACAGATC TTGTAATCAA TGCTCCTGAA GGAAAGGAAA CTTATGAAGG
 951 AACAATTAGC TTCTCAGGAC TATGCCTGGA TGATCATGAA GTTGTGCGG
 1001 AAAATCTTAC TTCCACAATC CTACAAGATG TCACATTAGC AGGAGGAACT
 1051 CTCTCTCTAT CGGATGGGGT TACCTTGCAA CTGCATTCTT TTAAGCAGGA
 1101 AGCAAGCTCT ACGCTTACTA TGCTCCAGG AACCACTCTG CTCTGCTCAG
 1151 GAGATGCTCG GGTTCAGAA CTGCACATCC TGATTGAAGA TACCGACAAC
 1201 TTTGTTCCCTG TAAGGATTCG CGCCGAGGAC AAGGATGCTC TTGTCTCATT
 1251 AGAAAACTT AAAGTTGCCT TTGAGGCTTA TTGGTCCGTC TATGACTTTC
 1301 CTCAATTTAA GGAAGCCTTT ACGATTCTCT TTCCTGAAC TCTAGGGCCT
 1351 TCTTTTGACA GTCTCTCCT AGGGGAGACC ACTTTGGAGA GAACCCAAGT
 1401 CACAACAGAG AATGACGCCG TTCGAGGTTT CTGGTCCCTA AGCTGGGAAG
 1451 AGTACCCCCC TTCTCTGGAT AAAGACAGAA GGATCACACC AACTAAGAAA
 1501 ACTGTTTCC TCACCTGGAA TCCTGAGATC ACTTCTACGC CATAA

45 The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E. coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 26A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 26B).

50 These experiments show that cp6735 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 27

The following *C. pneumoniae* protein (PID 4376784) was expressed <SEQ ID 53; cp6784>:

1 MNRRKARWVV ALFAMTALIS VGCCPWSQAK SRCSIDKYIP VVNRILLEVC
 51 LPEAENVEDL IESSAWVLT PEERFSGELV SICQVKDEHA FYNDSLHLHM
 101 TQAVPSYSAT YDCAVFGGP LPALRQRLDF LVREWQRGVR FKKIVFLCGE
 151 RGRYQSIIEQ EHFFDSRYNP FPTEENWESG NRVTPSSEEE IAKFVWMQML

201 LPRAWRDSTS GVRVTFLAK PEENRVVANR KDTLLLFPSY QEAFPGRVLF
 251 VSSQPFIGLD ACRVGQFFKG ESYDLAGPGF AQGVLYKHYA PRICLHTLAE
 301 WLKETNGCLN ISEGCFG*

A predicted signal peptide is highlighted.

5 The cp6784 nucleotide sequence <SEQ ID 54> is:

1 ATGAATAGAA GAAAAGCAAG ATGGGTTAGTG GCATTGTTTCG CAATGACGGC
 51 GCTCATTCTT GTTGGGTGTT GTCCCTTGGTC ACAAGCGAAA TCAAGATGTT
 101 CTATTGATAA GTATATTCCT GTAGTCAATC GTTTACTAGA AGTTTGTGGA
 151 CTTCCTGAAG CTGAGAAATGT TGAGGATTTA ATCGAGTCCT CGTCTGCTTG
 201 GGTACTGACT CCTGAAGAAC GTTTTCTGAG AGAGTTAGTC TCTATCTGTC
 251 AGGTAAAGA TGAGCATGCT TTCTATAACG ATTTGTCTTT ATTACATATG
 301 ACTCAGGCTG TGCCTTCGTA TTCTGCAACG TATGATTGTG CTGTAGTTTT
 351 TGGCGGGCCT TTGCCAGCGC TACGTCAGCG CTTAGATTTT TTGGTGCAG
 401 AGTGGCAGCG TGGCGTGCAG TTTAAGAAAA TCGTTTTTCT ATGTGGAGAG
 15 451 CGAGGGCGCT ATCAGTCTAT TGAAGAACAA GAGCATTTCT TTGATTCTCG
 501 GTACAATCCT TTCCCTACTG AAGAGAACTG GGAATCTGGT AACCAGAGTTA
 551 CTCCTCTTC TGAAGAAGAG ATTGCCAAAT TTGTTTGGAT GCAAATGCTT
 601 TTACCTAGAG CATGGCGAGA TAGTACTTCA GGAGTCAGAG TGACATTTCT
 651 TCTAGCAAAG CCAGAGGAAA ATCGTGTGGT TCGGAATCGT AAGGACACCT
 20 701 TACTTTTATT CCGTTCTTAT CAAGAAGCGT TTCCGGGACG CGTGTATTAT
 751 GTAAGTAGTC AACCCTTTAT CGGTTTAGAT GCTTGCAGGG TCGGGCAGTT
 801 TTCAAAGGG GAAAGCTATG ATCTTGCTGG ACCTGGATTG GCTCAAGGAG
 851 TCTTGAAGTA TCATTGGGCT CCAAGGATTT GTCTACATAC TTTAGCGGAA
 901 TGGTTAAAGG AAACGAACGG CTGCTTAAAT ATTTTCAGAG GTTGTTTTGG
 25 951 ATGA

The PSORT algorithm predicts a periplasmic location (0.894).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 27A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 27B). The GST-fusion product was used for FACS analysis (Figure 27C).

30 The cp6784 protein was also identified in the 2D-PAGE experiment (Cpn0498).

These experiments show that cp6784 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 28

The following *C.pneumoniae* protein (PID 4376960) was expressed <SEQ ID 55; cp6960>:

35 1 MNRRWNLVLA TVALALSVAS CDVRSKDKDK DQGSLEYKDK NKDPTNDIELS
 51 DNQKLSRTFG HLLARQLRKS EDMFFDIAEV AKGLQALVC KSAPLTETEFY
 101 EEKMAEVQKL VFEKKSKENL SLAEKFLKEN SKNAGVVEVQ PSKLQVKIHK
 151 EGAGKATSGK PSALLHYKGS FINGQVFSSS EGNNEPILLP LGQTIPGFAL
 201 GMQGMKEGET RVLYIHPDLA YGTAGQLPPN SLLIFEINLI QASADEVAAB
 40 251 PQEGNQGE*

A predicted signal peptide is highlighted.

The cp6960 nucleotide sequence <SEQ ID 56> is:

1 ATGAACAGAC GGTGGAATTT AGTTTTCAGCA ACAGTAGCTC TGGCACTCTC
 51 CGTCGCTTCT TGTGACGTAC GGTCTAAGGA TAAAGACAAG GATCAGGGGT
 45 101 CGTTAGTGGA ATATAAAGAT AACAAAGATA CCAATGACAT AGAATTATCC
 151 GATAATCAAA AGTTATCCAG AACATTTGGT CATTTATTAG CACGCCAATT
 201 ACGCAAGTCA GAAGATATGT TTTTGTGATAT TGCAGAAGTG GCTAAGGGGT
 251 TGCAGGCGGA ATTGGTTTGT AAAAGTGCTC CTTTAACAGA AACAGAGTAT
 301 GAAGAAAAAA TGGCTGAAGT ACAGAAGTTG GTTTTGTAAA AAAAATCAAA
 50 351 AGAAAATCTT TCATTGGCAG AAAAATCTTT AAAAGAAAAA AGCAAGAACG
 401 CTGGTGTTCG TGAAGTGCAA CCAAGTAAAT TGCAATACAA AATTATTAAA

5
 451 GAAGGTGCAG GAAAGCAAT TTCAGGTAAA CCTTCAGCTC TATTGCACTA
 501 CAAGGGTTCC TTCATCAATG GCCAAGTATT TAGCAGTTCA GAAGGCAACA
 551 ATGAGCCTAT CTTGCTTCCT CTAGGCCAAA CAATTCCTGG TTTTGTCTTA
 601 GGTATGCAGG GCATGAAAGA AGGAGAAACT CGAGTTCTCT ACATCCATCC
 651 TGATCTTGCT TACGGAACCG CAGGACAACT TCCTCCAAAC TCTTTATTAA
 701 TTTTGTAAAT TAACTTGATT CAGGCTTCAG CAGATGAAGT TGCTTGCTGA
 751 CCCCAGAAG GAAATCAAGG TGAATGA

The PSORT algorithm predicts periplasmic space location (0.930).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 28A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 28B) and for FACS analysis (Figure 28C).

The cp6960 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6960 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 29

The following *C.pneumoniae* protein (PID 4376968) was expressed <SEQ ID 57; cp6968>:

20
 1 **MKFLLYVPLL** LVLVSTG CDA KPVSEFPFSG KLSTQRFEPQ HSAEEYFSQG
 51 QEFLKKGNFR KALLCFG IIT HHFPRDILRN QAQYLIGVCY FTQDHPDLAD
 101 KAFASYLQLP DAEYSEELFQ MKYALAQRFA QGKRKRICRL EGFPKLMNAD
 151 BDALRIYDEI LTAFP SKDLG AQALYSKAAL LIVKNDL TEA TKTLKKLTLO
 201 FPLHLSSEA FVRLSEIYLQ QAKKEPHNLQ YLHFAKLNEE AMKKQHPNHP
 251 LNEVVSANVG AMREHYARGL YATGRFYEKK KKAEAAANIYY RTAITNYPDT
 301 LLVAKCQKRL DRISKHTS*

A predicted signal peptide is highlighted.

25 The cp6968 nucleotide sequence <SEQ ID 58> is:

30
 1 ATGAAATTTC TATTATACGT TCCACTTCTT CTTGTTCTCG TATCTACGGG
 51 GTGCGATGCA AAACCTGTTT CTTTTCAGCC CTTTTCAGGA AAGCTTTCCA
 101 CCCAGCGTTT TGAGCCTCAG CACTCTGCTG AAGAATATTT TTCTCAGGGA
 151 CAGGAATTCT TAAAAAAGG AAATTCAGA AAAGCTTAC TATGCTTTGG
 201 AATCATTTACG CATCACTTCC CTAGGGACAT CTTGCGTAAT CAAGCACAGT
 251 ATCTTATAGG AGTCTGTTAC TTCACGCAGG ATCACCAGG TTTAGCAGAC
 301 AAGGCATTTG CATCTTACTT ACAACTTCCT GATGCGGAGT ACTCTGAAGA
 351 GTTGTTCAG ATGAAATATG CGATTGCTCA AAGATTGCT CAAGGGAAGC
 401 GTAAACGGAT TTGTCGATTA GAGGGCTTCC CAAAATAAT GAATGCTGAT
 35 GAAGATGCGC TACGCATTTA TGACGAGATT CTAACAGCGT TTCCTAGTAA
 501 AGACTTAGGA GCTCAGGCC TCTATAGTAA AGCTGCGTTA CTTATTGTAA
 551 AAAACGATCT TACAGAAGCC ACCAAAACCT TAAAAAACT CACGTACAA
 601 TTTCTCTAC ATATTTTATC TTCAGAGGCC TTTGTACGTT TATCGGAAAT
 651 CTATTTACAG CAAGCTAAGA AAGAGCCTCA CAATCTCAA TATCTTCATT
 40 701 TTGCAAAGCT TAATGAAGAG GCAATGAAA AGCAGCATCC TAACCATCCT
 751 CTGAATGAGG TTGTTCTGCT TAATGTTGGA GCTATGCGGG AACATTATGC
 801 TCGAGGTTTG TATGCCACAG GTCGTTTCTA TGAGAAGAAG AAAAAAGCCG
 851 AGGCTGCGAA TATCTATTAC CGCACTGCGA TTACAAACTA CCCAGACACT
 901 TTATTAGTGG CTAAATGTCA AAAGCGTCTA GATAGAATAT CTAAGCATAC
 45 951 TTCCTAA

The PSORT algorithm predicts an inner membrane location (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 29A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 29B) and for FACS analysis (Figure 29C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6968 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 Example 30

The following *C.pneumoniae* protein (PID 4376998) was expressed <SEQ ID 59; cp6998>:

```

1  MKKLLRSALL SAAFAGSVGS LQALPVGNPS DPSLLIDGTI WEGAAGDPDCD
51 PCATWCDNIS LRAGFYGDYV FDRILKVDAP KTFMSGAKPT GSAAANYTTA
101 VDRENPAYNK HLHDAEWFTN AGFIALNIWD RFDVFTLGA SNGYIRGNST
151 AFNLVGLFGV KGTTVNANEL PNVLSNGVV ELYTDTSFWS SVGARGALWE
201 CGCATLGAEF QYAQSKPKVE ELNVICNVSQ FSVNKPQGYK GVAFPLPTDA
251 GVATATGTKS ATINYHEWQV GASLSYRLNS LVPYIGVQWS RATFDADNIR
301 IAQPKLPTAV LNLTAWNPSL LGNATALSTT DSFSDFMQIV SCQINKPKSR
351 KACGVTVGAT LVDADKWSLT AEARLINERA AHVSGQFRF*

```

15 A predicted signal peptide is highlighted.

The cp6998 nucleotide sequence <SEQ ID 60> is:

```

1  ATGAAAAAAC TCTTAAAGTC GCGCTTATTA TCCGCCGCAT TTGCTGGTTC
51 TGTGTGGCTCC TTACAAGCCT TGCCTGTAGG GAACCCCTTCT GATCCAAGCT
101 TATTAATGTA TGGTACAATA TGGGAAGGTG CTGCAGGAGA TCCTTGCGAT
151 CCTTGCCTTA CTGGTGCGA CGCTATTAGC TTACGTGCTG GATTTTACGG
201 AGACTATGTT TTGACCGGTA TCTTAAAGT AGATGCACCT AAAACATTTT
251 CTATGGGAGC CAAGCCTACT GGATCCGCTG CTGCAACTA TACTACTGCC
301 GTAGATAGAC CTAACCCGGC CTACAATAAG CATTTACACG ATGCAGAGTG
351 GTTCACATAA GCAGGCTTCA TTGCCTTAAA CATTTGGGAT CGCTTTGATG
401 TTTTCTGTAC TTTAGGAGCT TCTAATGGTT ACATTAGAGG AAACCTTACA
451 GCGTTCAATC TCGTTGGTTT ATTCGGAGTT AAAGGTACTA CTGTAATGTC
501 AAATGAACCT CCAACGTTT CTTTAAGTAA CGGAGTTGTT GAACTTTACA
551 CAGACACCTC TTCTCTTTGG AGCGTAGGCG CTCGTGGAGC CTTATGGGAA
601 TGCGGTTGTG CAACTTTGGG AGCTGAATTC CAATATGCAC AGTCCAAACC
651 TAAAGTTGAA GAACTTAATG TGATCTGTAA CGTATCGCAA TTCTCTGTAA
701 ACAAACCCAA GGGCTATAAA GGCGTTGCTT TCCCCTTGCC AACAGACGCT
751 GGCGTAGCAA CAGCTACTGG AACAAAGTCT GCGACCATCA ATTATCATGA
801 ATGGCAAGTA GGAGCCTCTC TATCTTACAG ACTAAACTCT TTAGTGCCAT
851 ACATTGGAGT ACAATGGTCT CGAGCAACTT TTGATGCTGA TAACATCCGC
901 ATTGCTCAGC CAAAACCTAC TACAGCTGTT TTAAACTTAA CTGCATGGAA
951 CCCTTCTTTA CTAGGAAATG CCACAGCATT GTCTACTACT GATTCGTTCT
1001 CAGACTTCAT GCAAAATTGT TCCTGTCAGA TCAACAAGTT TAAATCTAGA
1051 AAAGCTTGTG GAGTTACTGT AGGAGCTACT TTAGTTGATG CTGATAAATG
1101 GTCACTTACT GCAGAAGCTC GTTTAATTAA CGAGAGAGCT GCTCAGCTAT
1151 CTGGTCAGTT CAGATTCTAA

```

The PSORT algorithm predicts an outer membrane location (0.707).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 30A) and as a his-tag product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 30B) and for FACS analysis (Figure 30C).

45 The cp6998 protein was also identified in the 2D-PAGE experiment (Cpn0695) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6998 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 31

The following *C.pneumoniae* protein (PID 4377102) was expressed <SEQ ID 61; cp7102>:

```

1  MKHTFTKRVL FFFFLVIPIP LLLNLMVVG FFSFAAKANL VQVLHTRATN
51  LSIEFEKKLT IHKLFLDRLA NTLALKSYAS PSAEPYAQAY NEMMALSNITD
101 FSLCLIDPFD GSVRTKNPGD PFIRYLKQHP EMKKKLSAAV GKAFLLTIPG
151 KPLLHYLILV EDVASWDSTT TSGLLVSFYP MSFLQKDLFQ SLHITKGNIC
201 LVNKYGEVLF CAQDSESSFV FSLDLPLNPQ FQARSPSAIE IEKASGILGG
251 ENLITVSINK KRYLGLVLNK IPIQGTYTLS LVPVSDLIQS ALKVPLNICF
301 FYVLAFLLMW WIFSKINTKL NKPLQELTFC MEAAWRGNHN VRFEPQPYGY
10  351 EFNELGNIFN CTLLLLLNSI EKADIDYHSG EKLOKELGIL SSLQSALLSP
401 DFPTFPKVT FSSQHLRRRQL SGHFNQWTVQ DGGDTLLGII GLAGDIGLPS
451 YLYALSARSL FLAYASSDVS LQKISKDTAD SFSKTTEGNE AVVAMTFIKY
501 VEKDRSLELL SLSEGAPTMF LQRGESFVRL PLETHQALQP GDRLICLTGG
15  551 EDILKYFSQL PIEELLKDPL NPLNTEENLID SLTMMLNNET EHSADGTLTI
601 LSFS*

```

A predicted signal peptide is highlighted.

The cp7102 nucleotide sequence <SEQ ID 62> is:

```

1  ATGAACATA CCTTACCAA GCGTGTCTA TTTTTCCTT TTTTAGTGAT
51  TCCCATTCCTC CTACTCCTCA ATCTTATGGT CGTAGGTTTTT TTCTCATTTTT
101 CTGCCGCTAA AGCAAATTTA GTACAGGTCC TCCATACCCG TGCTACGAAC
151 TTAAGTATAG AATTCGAAAA AAACTGACG ATACACAAGC TTTTCTCGA
201 TAGACTTGCC AACACATTAG CCTTAAATC CTATGCATCT CCTCTGCAG
251 AGCCCTATGC ACAGGCATAC AATGAGATGA TGGCACTCTC CAATACAGAC
301 TTTTCCTTAT GCCTTATAGA TCCCTTTGAT GGATCTGTAA GGACGAAAAA
25  351 TCCTGGAGAC CCTTTCATTC GCTATCTAAA ACAGCATCCT GAAATGAAGA
401 AAAAGCTATC CGCAGCTGTA GGGAAAGCCT TTTTATTGAC CATTCAGGT
451 AAACCACTTT TACATTATCT TATTTAGTT GAAGATGTCG CATCTGGGA
501 TTCTACAACG ACTTCAGGAC TGCTTGTAAG TTTCTATCCC ATGCTCTTTT
30  551 TACAGAAAGA TTTATTCCAA TCCTTACACA TCACCAAAGG AAATATCTGC
601 CTTGTAAATA AGTATGGCGA GGTCTCTTC TGTCCTCAGG ACAGTGAATC
651 TTCTTTTGTA TTTCTCTAG ATCTCCCTAA TTTACGCAA TTCCAAGCAA
701 GAAGCCCCTC TGCCATAGAA ATTGAGAAAG CTTCTGGAAT TCTGGTGGG
751 GAGAACCATA TCACAGTGAG TATCAACAAG AAACGCTACC TAGGATTGGT
801 ACTGAATAAA ATTCCTATCC AAGGGACCTA CACTCTATCT TTAGTCCAG
35  851 TTTCTGATCT CATCCAATCC GCCTTGAAAG TTCCTCTCAA TATTGTGTTT
901 TTCTATGTAC TTGCTTTTCT CCTCATGTGG TGGATTTTCT CTAAGATCAA
951 CACCAAACTT AACAAAGCTC TTCAAGAACT GACCTTCTGT ATGGAAGCTG
1001 CCTGGCGAGG AAACATAAC GTGAGGTTTG AACCCAGGCC TTACGGTTAT
1051 GAATTCAATG AACTAGGAAA TATTTTCAAT TGCACTCTCC TACTCTTATT
40  1101 GAATTCCATT GAGAAAGCAG ATATCGATTA CCATTCAGGC GAAAAATTAC
1151 AAAAAGAAAT AGGGATTTTA TCTTCACTAC AAAGTGCCTT ACTAAGTCCG
1201 GATTTCCCTA CGTTCCTTAA AGTTACCTTT AGTTCCCAAC ATCTCCGGAG
1251 AAGGCAACTT TCCGGTCATT TTAATGGTTG GACAGTTCAA GATGGTGGCG
45  1301 ATACCTTTT AGGGATCATA GGGCTCGCTG GCGATATTGG TCTTCCTTCC
1351 TATCTCTATG CTTTATCCGC ACGGAGTCTT TTTCTTGCCCT ATGCTTCCTC
1401 GGACGTTTCG TTACAAAAAA TCAGCAAGGA TACTGCCGAC AGCTTCTCAA
1451 AAACAACAGA AGGCAATGAG GCTGTAGTTG CTATGACTTT CATTAAATAT
1501 GTAGAAAAAG ATCGATCTCT AGAGCTCCTC TCGTTAAGCG AGGGAGCTCC
50  1551 TACCATGTTT CTACAACGAG GAGAATCTTT CGTACGTCTC CCCTTAGAGA
1601 CTCACCAAGC TCTACAGCCT GGAGATCGGT TGATCTGCCT CACTGGAGGA
1651 GAAGACATCC TCAAGTACTT TTCTCAGCTT CCTATTGAAG AGCTCTTAAA
1701 AGATCCTTTA AACCCCTTAA ATACAGAGAA TCTTATGAT TCTCTAACCA
1751 TGATGTTAAA CAACGAAACC GAACATTCTG CAGATGGAAC TCTGACCATC
1801 CTTTCATTTT CATAA

```

55 The PSORT algorithm predicts an inner membrane location (0.338).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 31A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 31B).

These experiments show that cp7102 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 32

The following *C.pneumoniae* protein (PID 4377106) was expressed <SEQ ID 63; cp7106>:

```

5      1 MKDLGTLGGT SSTAKTVSPD GKVIMGRSQI ADGSWHAFMC HTDFSSNNVL
51     FDLNNTYKTL RENGRLNSI FNLQNMMLQR ASDHEFTFEG RSNIALGAGL
101    YVNALQNLPS NLAAQYFGIA YKIRPKYRLG VFLDHNFSH VPNNFNVSHN
151    RLWMGAFIGW QSDALGSSV KVSFGYKQK ATITREQLN TEAGSGESHF
201    EGVAAQIEGR YGKSLGGHVR VQPFLGLQFV HITRKEYTEN AVQFPVHYDP
10     251 IDYSTGVVVL GIGSHIALVD SLHVGTTRMG EQNFAAHTDR FSGSIASIGN
301    FVFEKLDVTH TRAFEMRVN YELPYLQSLN LILRVNQQL QGVMGFSSDL
351    RYALGF*

```

The cp7106 nucleotide sequence <SEQ ID 64> is:

```

15      1 ATGAAAGATT TGGGGACTCT TGGGGGTACC TCTTCTACAG CAAAAACAGT
51     GTCCCCAGAT GGTAAAGTGA TCATGGGTAG ATCACAATT GCTGATGGCA
101    GTTGGCACGC ATTTATGCT CATACGGATT TCTCCTCTAA TAATGTACTC
151    TTTGATCTCG ATAATACGTA TAAACTCTA AGAGAAAATG GCCGTCAGCT
201    AAATTCATA TTCAACCTAC AAAATATGAT GTTACAGAGA GCCTCAGATC
20     251 ATGAGTTCAC AGAGTTTGA AGGAGTAACA TCGCTCTTGG TGCCGGGCTT
301    TATGTGAATG CCTTGCAGAA TCTCCCTAGC AATTTAGCAG CACAATATTT
351    TGAATCGCA TACAAAATAC GTCCTAAATA TCGTTTGGGG GTGTTTTTGG
401    ACCATAATTT CAGCTCCAC GTTCCTAATA ATTTTAACGT AAGCCACAAT
451    AGACTCTGGA TGGGAGCCTT TATGGATGG CAGGATCTG ATGCTCTAGG
501    ATCTAGTGTC AAGGTGTCTT TCGGATATGG AAAACAAAAA GCCACGATTA
25     551 CAAGAGAGCA ATTAGAGAT ACAGAAGCG GGAGTGGGA GAGCCATTTT
601    GAAGGGGTCG CTGCTCAGAT AGAAGGGCGG TATGGTAAGA GCCTCGGAGG
651    ACATGTCAGG GTCCAGCCTT TCCTAGGACT GCAGTTTGTC CACATTACAA
701    GGAAAGAATA TACCGAAAAT GCAGTGCAAT TTCCTGTACA CTATGATCCT
751    ATAGACTATT CTACAGGTGT AGTGATTTA GGAATTGGAT CTCATATTGC
30     801 ACTGTAGAT TCTTTACATG TAGGCACACG CATGGGAATG GAGCAAACT
851    TTGCAGCCCA TACGACAGG TTCTCAGGAT CTATAGCGTC TATTGGAAC
901    TTTGTGTTTG AAAAGCTTGA TGTGACTCAC ACAAGGCAT TTGCGGAAT
951    GCGTGTC AAC TATGAGCTTC CCTATCTACA GTCTCTGAAT CTTATTCTAC
1001   GAGTTAATCA ACAGCCTCTA CAAGGGGTTA TGGGATTTTC CAGTGATCTT
35     1051 AGGTATGCCT TAGGATTCTA A

```

The PSORT algorithm predicts a cytoplasmic location (0.224).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 32A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 32B) and for FACS analysis (Figure 32C).

This protein also showed very good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7106 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

45 Example 33

The following *C.pneumoniae* protein (PID 4377228) was expressed <SEQ ID 65; cp7228>:

```

1      1 MTAVLILTSE PSEESARSLA RHLITERLAS CVHVFPKGTS TYLWEGKLCE
51     SEEHHIQIKS IDIRFSEICL AIQEFSGYEV PEVLLFFPIEN GDFRYLNLWT
101    ILSYPEKPPL SD*

```

The cp7228 nucleotide sequence <SEQ ID 66> is:

```

1  ATGACTGCTG TTCTTATTCT TACATCTTTC CCTTCGGAGG AAAGTGCTCG
51 CTCCCTAGCT AGACATCTGA TTACAGAGCG TCTTGCTTCC TGTGTGCATG
101 TATCCCTAA AGGCACATCG ACATATCTAT GGAAGGCAA GCTATGTGAG
151 TCTGAAGAAC ATCATATACA AATCAAATCG ATAGACATAC GCTTCTCGGA
201 AATTTGTCTT GCTATTCAGG AGTTCTCTGG CTATGAGGTT CCTGAAGTCT
251 TACTATTTC TATTGAAAAT GGGGATCCGA GGTACTTGAA TTGGTTAACG
301 ATTCTCAGCT ATCCAGAGAA GCCTCCGCTT TCAGATTAG

```

The PSORT algorithm predicts an inner membrane location (0.040).

- 10 The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 33A (his-tag = left-hand arrow, GST = right-hand arrow). The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 33B) and FACS analysis.

These experiments show that cp7228 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 34

The following *C.pneumoniae* protein (PID 4377170) was expressed <SEQ ID 67; cp7170>:

```

1  MNSKMLKHLR LATLSFSMFF GIVSSPAVYA LGAGNPAAFPV LPGVNPEQTG
51 WCAFQLCNSY DLFAALAGSL KFGFYGDYVF SESAHITNVP VITSVTTSSTG
101 GTPPTITSTT KNVDFDLNNS SISSSCVFAT IALQETSPAA IPLLDIAFTA
151 RVGGLKQYR LPLNAYRDFT SNPLNAESEV TDGLIEVQSD YGIVWGLSLQ
201 KVLWKDGVSF VGVSADYRHG SSPINYIIVY NKANPEIYFD ATDGNLSYKE
251 WSASIGISTY LNDYVLPYAS VSIGNTSRKA PSDSFTELEK QFTNFKFKIR
301 KITNFDVRNF CFGTTCCISN NFYYSVEGRW GYQRAINITS GLQF*

```

A predicted signal peptide is highlighted.

- 25 The cp7170 nucleotide sequence <SEQ ID 68> is:

```

1  ATGAATAGCA AGATGCTAAA ACATTTACGT TTAGCAACCC TTCTCTTCTC
51 TATGTTCTTC GGGATTGTAT CTCTCCCGC AGTATATGCC CTAGGGGCTG
101 GAAACCCTGC AGCTCCAGTA CTCCAGGTG TGAATCCTGA GCAAACGGGA
151 TGGTGTGCCT TCCAACCTTG TAATAGTTAC GATCTTTTTC CTGCTCTTGC
201 AGGAAGCCTC AAATTTGGGT TCTATGGAGA TTATGTCTTC TCAGAAAGTG
251 CCCATATTAC CAATGTCCCT GTCATTACCT CCGTTACGAC TTCAGGCACA
301 GGAACAACGC CAACCATTAC CTCTACAAC AAAAAAGTAG ACTTTGATCT
351 TAACAACAGC TCCATCAGCT CGAGCTGTGT TTTTGCAACC ATAGCTCTAC
401 AGGAAACATC CCCAGCTGCC ATTCCCCTTT TAGATATAGC CTTCACTGCA
35 CGTGTCCGAG GACTTAAGCA GTACTACCGC CTCCCTCTCA ATGCTTACAG
501 AGACTTCACT TCAAATCCTT TAAATGCAGA ATCTGAAGTT ACAGATGGTC
551 TCATTGAAGT CCAGTCAGAC TATGGAATTG TCTGGGGTCT GAGTTTACAA
601 AAAGTATTGT GGAAAGATGG AGTGTCTTTT GTAGGGGTGA GCGCTGACTA
651 CCGTCACGGT TCCAGTCCCA TCAACTATAT CATCGTTTAC AACAAGGCCA
40 ACCCCGAGAT CTATTTCGAT GCTACTGATG GAAACCTAAG CTATAAGAA
751 TGGTCTGCAA GCATCGGCAT CTCTACGTAT CTTAATGACT ATGTGCTTCC
801 CTATGCATCC GTATCTATAG GAAATACTTC AAGAAAAGCT CCTTCTGATA
851 GCTTCACAGA ACTCGAAAAG CAATTTACGA ATTTTAAATT TAAAATTTCGT
901 AAAATCACAA ACTTCGACAG AGTAAACTTC TGCTTCGAA CTACCTGCTG
45 951 CATCTCAAAT AACTTCTACT ATAGTGTAGA AGGCCGTTGG GGATATCAGC
1001 GTGCTATCAA CATTACGTCA GGTCTGCAGT TTTAG

```

The PSORT algorithm predicts a bacterial outer membrane location (0.936).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product.

The purified GST-fusion product is shown in Figure 34A. The GST-fusion protein was used to

- 50 immunise mice, whose sera were used in a Western blot (34B) and for FACS analysis (34C).

The cp7170 protein was also identified in the 2D-PAGE experiment (Cpn0854).

These experiments show that cp7170 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 35

5 The following *C.pneumoniae* protein (PID 4377072) was expressed <SEQ ID 69; cp7072>:

```

1  MDIKKLFCLF LCSSLIAMSP IYKGTGDYK LTLTGINIID RNGLSETICS
51  KEKLLKYYTKV DFLAPQPYQK VMRMKYNKR DNVSCLTAYH TNGQIKQYLE
101 CLNNRAYGRY REWHVNGNIK IQAEVIGGIA DLHPSAESGW LFDQTTFFAYN
151 DEGILEAAIV YEKGLLEGSS VYYHTNGNIW KECVYHKGVP QGKFLTYTSS
10  201 GKLLKEQNYQ QGKRHGLSIR YSEDSEEDVL AWEEYHEGRL LKAEYLDPOQT
251 HEIYATIHES NGIQAIYGKY AVIETRAFVR GEPYGVTRF DNSGTQIVQT
301 YNLLQGAHKG EEEFFYPETG KPCLLLNWHE GILNGIVKTW YPGGTLESCK
351 ELVNNKKSGL LTIYYPEGQI MATEEYDNDL LIKGEYFRPG DRHPYSKIDR
401 GCGTAVFFSS AGTITKKIPY QDGKPLLN*

```

15 A predicted signal peptide is highlighted.

The cp7072 nucleotide sequence <SEQ ID 70> is:

```

1  ATGGATATAA AAAAATCTTT TTGCTTATTT CTATGTTCTT CTCTAATTGC
51  CATGAGTCCC ATTTATGGGA AAACAGGTGA CTATGAGAAA CTCACCCCTTA
101 CAGGGATCAA TATCATTTGAT AGAAACGGCC TGTCAGAAAC TATTTGCTCT
20  151 AAAGAGAAGC TAAAGAAATA CACCAAGGTA GACTTCTCTG CTCCCCAGCC
201 CTATCAAAAG GTCATGAGGA TGTATAAAAA CAAACGCGGA GATAACGTTT
251 CTTGTTTAAC AGCCTATCAC ACTAACGGGC AAATTAAGCA GTACCTGGAG
301 TGTCTCAATA ATCGTGCTTA TGGAAGATAT CGTGAATGGC ACGTCAACGG
351 GAATATCAAA ATCCAAGCTG AGGTATTCGG AGGTATTGCG GATCTTCATC
25  401 CCTCAGCAGA GTCTGGCTGG CTATTTGATC AAACACATT TGCCATAAAT
451 GATGAAGGTA TCTTAGAAGC CGCTATCGTC TATGAAAAAG GCGTGCCTCGA
501 AGGATCTTCG GTGTATTACC ATACTAATGG GAATATTTGG AAAGAGTGTC
551 CCTATCATAA GGGAGTTCCCT CAAGGTAAAT TCCTGACATA CACATCTTCG
601 GGGAAACTGC TCAAAGAACA GAATTACCAA CAAGGCAAAA GACACGGTCT
30  651 TTCGATTTCG TACAGCGAAG ATTCCGAAGA AGATGTTTTA GCCTGGGAAG
701 AATATCATGA GGGACGACTC CTAAAGCAG AGTACTTAGA TCCTCAAAC
751 CACGAAATCT ATGCGACTAT ACACGAAGGG AACGGCATTC AAGCAATCTA
801 CGGCAAGTAT GCCGTTATAG AAAC TAGGGC ATTTTACCGA GGGGAACCTT
851 ATGGAAGAGT TACCAGATTC GACAACTCCG GAACACAGAT TGTCCAAACG
35  901 TATAACCTTT TGCAAGGCGC GAAGCACGGA GAAGAATTTT TCTTTTATCC
951 TGAGACAGGG AAACCCAAGC TGCCTCTTAA TTGGCATGAA GGAATTTTAA
1001 ATGGGATAGT AAAAATCTGG TATCCCGGAG GAACCTTAGA AAGTTGTAAA
1051 GAACTCGTAA ATAACAAAA ATCCGGGTTA CTGACCATT ACTACCCTGA
1101 AGGACAGATC ATGGCGACCG AAGAGTATGA TAATGATCTT CTAATTAAAG
40  1151 GAGAGTACTT CCGCCCTGGA GACCGTCATC CCTACTCTAA AATAGATCGT
1201 GGTGTGGGGA CTGCAGTATT TTTCTCGTCG GCGGGAAC TA TTAATAAAAA
1251 AATCCCCAT CAGGACGGCA AACCTTTGCT CAAC TAG

```

The PSORT algorithm predicts a periplasmic location (0.688).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 35A) and as a GST-
45 fusion product (Figure 35B). The recombinant his-tag protein was used to immunise mice,
whose sera were used in a Western blot (Figure 35C) and for FACS analysis.

These experiments show that cp7072 is a useful immunogen. These properties are not evident from the sequence alone.

Example 36

50 The following *C.pneumoniae* protein (PID 4376879) was expressed <SEQ ID 71; cp6879>:

1 MATPAQKSPT FQDPSFVREL GSNHPVFSPL TLEERGEMAI ARVQCCGWNH
 51 TIVKVSLLIL ALLTILGGGL LVGLLPAVPM FIGTGLIALG AVIFALALIL
 101 CLYDSQGLPE ELPPVPEPQQ IQIEDLRNET REVLEGTLLLE VLLKDRDAKD
 151 PAVPQVVVDC EKRLGMLDRK LRREEEILYR STAHLKDEER YEFLLLELLE
 201 RSLVADRLEF NRRSYERFVQ GIMTVRSEEG EKEISRLQDL ISLQQQTVQD
 251 LRSRIDDEQK RCWTALQRIN QSQKDIQRAH DREASQRACE GTEMDCARERQ
 301 QLEKDLRRQL KSMQEWIEMR GTIHQQEKAW RKQNAKLERL QEDLRLTGIA
 351 FDEQSLFYRE YKEKYLSQL DMQKILQEVN AEKSEKACLE SLVHDYEQKL
 401 EQKDANLKKK AAVWEEELGK QQQEDYEQTQ EIRRLSTFIL EYQDSLREAE
 451 KVEKDFQELQ QRYSRLOEEK QVKEKILEES MNHFADLFK AQKENMAYKK
 501 KLDLEGAAGA PTEIGEDDDW VLTDSASLSQ KKIRELVEEN QELLKALAFK
 551 SNELTQLVAD AVEAEKEISK LREHIEBQKE GLRALDKMHA QAIKDCEAAQ
 601 RKCCDLESLL SPVREDAGMR FELEVELQRL QEENALRAE VERLEQEQQFQ
 651 G*

15 The cp6879 nucleotide sequence <SEQ ID 72> is:

1 ATGGCAACAC CCGCTCAAAA ATCCCTTACA TTTCAAGATC CTAGTTTTGT
 51 AAGAGAGCTA GGCAGTAACC ACCCTGTCTT TTCCCGCTA ACGCTTGAGG
 101 AAAGAGGGGA GATGGCAATA GCTCGAGTCC AGCAGTGTGG ATGGAATCAT
 151 ACAATTGTTA AGGTAAGTCT TATTATTCTT GCTCTTCTTA CTATTTTAGG
 201 GGGAGGATTA CTCGTAGGAT TGCTGCCAGC AGTTCTATAG TTTATTGGAA
 251 CAGGTCTGAT TGCTTTGGGA GCCGTATAT TGTCTTGGC TTTGATTTTA
 301 TGTCTTTATG ATTCTCAGGG CCTTCCTGAG GAACTCCCTC CGGTTCCTGA
 351 ACCACAACAA ATTCTAGATTG AAGATTAAAG AAACGAGACC AGAGAAGTTC
 401 TTGAAGGGAC TCTTTTAGAG GTTCTCTTAA AGGATAGAGA CGCTAAGGAC
 451 CCTGCGGTGC CCCAGGTGGT TGTAGACTGT GAAAGCGTCT TTGGAATGTT
 501 GGATCGTAAG CTGCGACGTG AAGAGGAGAT TCTGTATCGC TCGACGGCCC
 551 ATCTTAAAGA CGAGGAAAGG TATGAGTTCT TGCTGGAGCT CTTGGAAATG
 601 CGTAGTCTGG TTGCCGATCG GCTAGAATT TAAACCGTAGAA GTTATGAGCG
 651 ATTTGTTCAA GGAATTATGA CAGTTAGATC AGAGGAGGGG GAAAAAGAGA
 701 TTTCTCGTCT ACAAGATCTA ATCAGTTTGC AGCAGCAGAC GGTGCAAGAT
 751 TTAAGGAGTC GGATCGATGA CGAGCAGAAG AGATGCTGGA CGGCTTTACA
 801 ACGTATTAAAC CAATCTCAGA AGGATATACA ACGGGCTCAT GATCGCGAGG
 851 CTTCGCAGCG TGCTGTGAG GGCACAGAGA TGGATTGTGC AGAACGCCAG
 901 CAACTGGAGA AGGATTTAAG GAGACAGCTG AAATCTATGC AGGAGTGGAT
 951 TGAGATGAGG GGCACAATCC ATCAACAAGA GAAGGCTTGG CGTAAGCAGA
 1001 ATGCCAAATT AGAAAGATTA CAAGAGGATC TGAGACTTAC TGGGATTGCT
 1051 TTTGACGAAC AATCTCTGTT CTATCGCGAA TATAAAGAGA AATATCTGAG
 1101 TCAGAACTA GATATGCAAA AGATTTTACA GGAAGTCAAC GCAGAGAAA
 1151 GTGAGAAAGC TTGCTTAGAG AGTCTGGTCC ATGACTATGA GAAGCAGCTC
 1201 GAACAAAAG ATGCTAATCT GAAGAAAGCA GCAGCTGTTT GGGAGGAAGA
 1251 ATTAGGGAAG CAGCAACAGG AAGACTACGA ACAAAACCAA GAAATTAGAC
 1301 GTCTGAGTAC ATTCATTCTT GAGTACCAGG ACAGTCTGCG TGAGGCAGAA
 1351 AAAGTTGAGA AAGATTTCCA AGAGCTACAA CAAAGGTATA GCCGCTTTCA
 1401 AGAGGAGAAA CAGGTAAAAG AAAAAATCTT AGAAGAAAGT ATGAATCATT
 1451 TTGCCGATCT CTTTGAGAAG GCTCAAAAGG AAAACATGGC CTACAAGAAG
 1501 AAGTTAGCGG ATTTAGAGGG TGCCGCTGCT CTTACTGAGA TCGGTGAGGA
 1551 CGATGACTGG GTACTCACAG ATTCTGCTTC TCTCAGCCAG AAGAAGATCC
 1601 GCGAACTCGT GGAAGAGAAT CAAGAACTCC TGAAAGCACT TGCATTTTAA
 1651 TCTAACGAAT TGACTCAACT GGTGCGGAT GCTGTAGAAG CTGAAAAAGA
 1701 AATCAGCAAG CTTCGAGAAC ACATAGAAGA GCAGAAAGAA GGATTACGAG
 1751 CTCTTGATAA GATGCATGCA CAAGCGATCA AAGATTGCGA AGCTGCTCAG
 1801 AGAAAAATGCT GTGACCTTGA GAGCCTTCTC TCTCCTGTTC GAGAAGATGC
 1851 TGGAATGAGA TTTGAGCTAG AGGTCGAGCT TCAAGATTTG CAAGAAGAAA
 1901 ATGCACAGCT TAGAGCGGAG GTTGAAAGAC TAGAGCAAGA GCAATTTCAA
 1951 GGATAA

The PSORT algorithm predicts an inner membrane location (0.646).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 36A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 36B) and for FACS analysis.

60 These experiments show that cp6879 is useful immunogen. These properties are not evident from the sequence alone.

Example 37

The following *C.pneumoniae* protein (PID 4376767) was expressed <SEQ ID 73; cp6767>:

```

1 MIKQIGRFFR AFIFIMPLSL TSCESKIDRN RIWIVGTNAT YPPFEYVDAQ
51 GEVVGFDIDL AKAISEKLKQ QLEVREFAFD ALILNLKKHR IDAILAGMSI
101 TFSRQKEIAL LPYYGDEVQE LMVVSQRSLE TPVLPLTQYS SVAVQGTTFQ
151 EHYLLSQPGI CVRSFDSLE VIMEVRYGKS PVAVLEPSVG RVVLKDFPNL
201 VATRLELPPE CWVLGCGLGK AKDRPEEIQT IQQAITDLKS EGVISLTKK
251 WQLSEVAYE*

```

The cp6767 nucleotide sequence <SEQ ID 74> is:

```

10 1 ATGATAAAAC AAATAGGCCG TTTTTTTAGA GCATTATTT TTATAATGCC
51 TTTATCTTTA ACAAGTTGTG AGTCTAAAAT CGATCGAAAT CGCATCTGGA
101 TTGTAGGTAC GAATGCTACA TATCCTCCTT TTGAGTATGT GGATGCTCAG
151 GGGGAAGTTG TAGGTTTCGA TATAGATTG GCAAAGGCAA TTAGTGAAAA
201 ACTTGGAAG CAATTGGAAG TTAGAGAATT CGCTTTCGAT GCTTTAATTT
15 251 TAAATTTAAA AAAACATCGT ATCGATGCAA TTTTAGCAGG AATGTCCATT
301 ACTCCTTCGC GTCAGAAGGA AATCGCCCTG CTCCCTATT ATGGCGATGA
351 GGTTCAGAG CTGATGGTGG TTTCTAAGCG GTCTTTAGAG ACCCCTGTGC
401 TTCCCTAAC ACAGTATCTC TCTGTTGCTG TTCAGACAGG AACGTTTCAG
451 GAGCATTATC TTTATCTCA GCCCGGAATT TGTGTCCGTT CTTTGTATAG
20 501 CACCTTGGAG GTGATTATGG AAGTTCGTTA TGGGAAATCT CCGGTTGCCG
551 TTCTAGAACC CTCGGTAGGA CGTGTCTGTC TTAAAGACTT CCCTAATCTT
601 GTTGCAACAA GATTAGAGCT CCCTCCTGAA TGTGGGTGT TGGGCTGTGG
651 TCTCGGCGTA GCTAAAGATC GTCCTGAAGA AATACAAACG ATTCAACAAG
701 CGATTACAGA TTAAAGAGC GAAGGGTGA TTCAATCTTT AACCAAGAAA
25 751 TGGCAACTTT CTGAAGTTC TTACGAATAG

```

The PSORT algorithm predicts an inner membrane location (0.083).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified his-tag product is shown in Figure 37A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 37B) and for FACS analysis (Figure 37C). The GST-fusion was also used in a Western blot (Figure 37D).

The cp6767 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6767 is a useful immunogen. These properties are not evident from the sequence alone.

Example 38

The following *C.pneumoniae* protein (PID 4376717) was expressed <SEQ ID 75; cp6717>:

```

1 MMSRLRFRLA ALGIFFILLV PNSVSAKTIV ASDKEKVGVL VYDNSVEAFQ
51 QILDCIDHAN FYVELCPCMT GGRTLKEMVD HLEARMDLVP ELCSYIIIQP
101 TFTDAEDQKL LKALKERHPN RFFYVFTGCP PSTSILAPNV IEMHIKLSII
40 151 DGKYCILGTT NFEEFMCTPG DEVPEKVDNP RLFVSGVRRP LAFRDQDIML
201 RSTAFGLQLR EBYHKQFAMW DYYAHMMWFI DNPEQFAGAC PPLTLEQAE
251 TVFPGFDKHE DLVLVDSSKI RIVLGGPHDK QPNPVTQEYL KLIQGARSSV
301 KLAHMYFIPK DELLNALVDV SHNHGVHLSL ITNGCHELSP AITGPIYAWN
351 RINYFALLYG KRYPLWKKWF CEKLPYERV SIYEFWIWET QLHKKCMID
45 401 DEIFVIGSYN FGKSDAFDY ESIVVIESPE VAAKANKVFN KDIGLSIPVS
451 HGDIFSWYFH SVHHTLGHQLQ LTYMPA*

```

A predicted signal peptide is highlighted.

The cp6717 nucleotide sequence <SEQ ID 76> is:

1 ATGATGAGTC GGTTCGCTTT TCGCTTGGCA GCTCTTGGAA TATTTTATAT
 51 TTTGCTGGTT CCTAATTCCTG TTTTACGAAA GACAAATCGTA GCTTCAGACA
 101 AGGAGAAGGT TGGAGTTCTT GTTTATGACA ATAGTGTAGA GGCCTTTCAA
 151 CAGATATTGG ATTGCATAGA TCATGCAAAAT TTTTATGTAG AACTGTGTCC
 201 CTGCATGACA GGAGGCCGAA CGCTTAAAGA GATGGTAGAT CACCTCGAGG
 251 CTCGTATGGA TCTGGTTCCA GAGCTCTGTA GCTATATCAT TATCCAACCC
 301 ACGTTTACCG ATGCTGAAGA CCAAAAATTA CTCAAAGCTC TCAAAGAACG
 351 TCATCCCAAC CGGTTTTTCT ACGTTTTTAC AGGGTGCCCA CCCTCAACAA
 401 GCATCCTCGC TCCTAATGTC ATTGAAATGC ATATCAAAC TCTATCATC
 451 GATGGGAAAT ATTGTATTTT AGGTGGTACC AATTTTGAAG AGTTTATGTG
 501 CACTCCAGGG GATGAGGTTT CTGAGAAAGT GGATAACCCA CGTTTATTTG
 551 TCAGTGGAGT GCGTCGGCCC CTAGCATTTT GTGATCAGGA TATCATGTTG
 601 CGTTCTACAG CATTTCGGTTT GCAGCTCAGA GAAGAATATC ATAAGCAATT
 651 TGCTATGTGG GACTACTATG CACATCATAT GTGGTTCATT GATAATCCTG
 701 AACAGTTTGC AGGCGCCTGT CCTCCACTGA CTTTAGAACA AGCCGAGGAG
 751 ACAGTATTTT CTGGATTTGA CAAACATGAA GATCTTGTTC TTGTCGACTC
 801 TTCCAAGATC AGGATAGTTT TAGGTGGTCC CCACGATAAG CAACCCAATC
 851 CTGTGACTCA AGAATATTTG AAACCTATCC AGGGAGCTAG ATCTTCTGTG
 901 AAGCTTGCTC ACATGTATTT CATCCCTAAG GACGAGCTTT TAAATGCTCT
 951 TGTCGACGTT TCTCATAATC ACGGTGTTCA TCTGAGTTTA ATTACGAACG
 1001 GCTGTCATGA ATTAAGTCCT GCAATTACAG GACCCTATGC TTGGGGAAAC
 1051 CGTATTAACT ATTTTCGCCTT GCTCTATGGG AAACGGTATC CTCCTTGGAA
 1101 AAAATGGTTT TCGGAAAAGC TAAAACCTTA TGAGCGGGTT TCTATTTATG
 1151 AGTTTGCTAT TTGGGAAACG CAGTTGCACA AGAAGTGTAT GATTATCGAT
 1201 GATGAAATTT TTGTGATCGG AAGTTATAAT TTTGGAAAGA AAAGTGATGC
 1251 CTTTGATTTAC GAAAGTATTG TAGTTATCGA ATCTCCAGAA GTCGCTGCAA
 1301 AAGCTAACAA AGTCTTCAAT AAAGATATCG GATTGTGAT TCTGTAAAGT
 1351 CATGGCGACA TTTTCTCTTG GTATTTCCAT TCCGTACACC ACACCTTTGGG
 1401 ACATTTGCAG CTGACCTATA TGCCAGCCTA G

30 The PSORT algorithm predicts a periplasmic location (0.939).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 38A), as a his-tagged protein, and as a GST/his fusion product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 38B) and for FACS analysis.

35 These experiments show that cp6717 is a useful immunogen. These properties are not evident from the sequence alone.

Example 39

The following *C.pneumoniae* protein (PID 4376577) was expressed <SEQ ID 77; cp6577>:

1 MKKLLFSTFL LVLGSTSAAH ANLGYVNLKR CLEESDLGKK ETEELEAMQO
 51 QFVKNAEKIE EELTSIYNKL QDEDYMESLS DSASEELRKK FEDLSGEYNA
 101 YQSQQYQSIN QSNVKRIQKL IQEVKIAAES VRSKEKLEAI LNEEAVLAI A
 151 PGTDKTTEII AILNESFKKQ N*

A predicted signal peptide is highlighted.

The cp6577 nucleotide sequence <SEQ ID 78> is:

1 ATGAAAAAAT TATTATTTTC TACATTTCTT CTTGTTTTAG GATCAACAAG
 45 51 CGCAGCTCAT GCAAAATTTAG GCTATGTTAA TTTAAAGCGA TGTCTTGAAG
 101 AATCCGATCT AGGTAAAAAG GAAACTGAAG AATTGGAAGC TATGAAACAG
 151 CAGTTTGTA AAAATGCTGA GAAATAGAA GAAGAACTCA CTTCTATTTA
 201 TAATAAGTTG CAAGATGAAG ATTACATGGA AAGCCTATCG GATTCTGCCT
 251 CTGAAGAGTT GCGAAAGAAA TTCGAAGATC TTTTCAAGAG GTACAAATGCG
 50 301 TACCAGTCTC AGTACTATCA ATCTATCAAT CAAAGTAATG TAAAACGCAT
 351 TCAAAAACCTC ATTCAAGAAG TAAAAATAGC TGCAGAAATCA GTGCGGTCCA
 401 AAGAAAAACTC AGAAGCTATC CTTAATGAAG AAGCTGTCTT AGCAATAGCA
 451 CCTGGGACTG ATAAACAAC CGAAATTATT GCTATTCTTA ACGAATCTTT
 501 CAAAAACAA AACTAG

55 The PSORT algorithm predicts a periplasmic space location (0.932).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 39A) and as a GST-fusion product (Figure 39B). The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 39C) and for FACS analysis.

The cp6577 protein was also identified in the 2D-PAGE experiment.

- 5 These experiments show that cp6577 is a useful immunogen. These properties are not evident from the sequence alone.

Example 40

The following *C.pneumoniae* protein (PID 4376446) was expressed <SEQ ID 79; cp6446>:

```

10      1  MKQPMSLIFS SVCLGLGLGS LSSCNQKPSW NYHNTSTSEE FFVHGKNSVS
      51  QLFHYPSAFR TTQIFSEEHN DPYVVAKTDE ESRKIWREIH KNLKIKGSYI
     101  PISTYGLMH  PKSAALTLKT YRPHPIWING YERSFNIDTG KYLKNISRRL
     151  TSHDGPKNRA VLNLIKSSGR RCNAIGLEMT EEDFVIARRR EGVVSLYPVE
     201  VCSYPQGNPF VIAYAWIADE SACSEVLPV  KGYVSLVWES VSSDSLNAF
     251  GDSFAEDYLR STFLANGTSI LCVHESYKKV PPQP*
```

- 15 A predicted signal peptide is highlighted.

The cp6446 nucleotide sequence <SEQ ID 80> is:

```

      1  ATGAAACAGC CCATGTCTCT TATCTTTTCA AGTGTATGTT TAGGATTAGG
      51  TCTTGGATCT CTTTCCTCCT GTAATCAAAA GCCCTCTTGG AATTATCACA
     101  ACACTTCAAC GAGCGAAGAA TTCTTTGTTC ATGGAAATAA GAGTGTTCG
     151  CAACTGCCTC ATTATCCTTC TGCATTTCGT ACGACTCAAA TCTTTTCTGA
     201  AGAGCACAAT GATCCTTATG TCGTAGCTAA GACTGATGAA GAGTCTCGTA
     251  AAATTTGGAG AGAAATCCAT AAAAATCTCA AAATCAAAGG TTCTTACATT
     301  CCCATATCGA CTTATGGAAG TCTGATGCAC CCAAAATCAG CAGCTCTTAC
     351  ATTA AAAACG TATCGTCCAC ATCCTATTTG GATAAATGGA TACGAGCGTT
     401  CTTTAAATAT AGACACAGGA AAGTACTTAA AAAACGGAAG TCGCCGTAGA
     451  ACTTCTCACG ATGGTCCGAA AAATCGAGCT GACTGAATC TCATTAAATC
     501  TTCGGGACGA CGCTGTAATG CTATAGGCCCT TGAGATGACA GAAGAAGACT
     551  TTGTAATAGC TAGAAGGCCGA GAAGGTGTTT ATAGCCTGTA TCCCGTTGAA
     601  GTGTGCTCGT ATCCTCAGGG GAATCCTTTT GTCATTGCTT ATGCCGTGGAT
     651  TGCAGATGAG AGTGCCTTGT CAAAAGAGGT CCTACCTGTA AAAGGGTACT
     701  ATTCTTTAGT CTGGGAAAGC GTTCTCTTCT CTGATTCTCT GAATGCTTTT
     751  GGAGATTCCCT TTGCAGAGGA CTACCTCAGA AGCACGTTTT TAGCAAACGG
     801  AACTTCTATA CTCTGTGTTC ATGAAAGCTA TAAGAAAGTT CCTCTCAGC
     851  CCTAA
```

- 35 The PSORT algorithm predicts an inner membrane location (0.177).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion product. The GST-fusion product is shown in Figure 40A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 40B) and for FACS analysis.

- 40 These experiments show that cp6446 is a useful immunogen. These properties are not evident from the sequence alone.

Example 41

The following *C.pneumoniae* protein (PID 4377108) was expressed <SEQ ID 81; cp7108>:

```

45      1  MSKKIKVLGH LTLCTLFRGV LCAALSNIG YASTSQESPV QKSIEDWKGY
      51  TFTDLELLSK EGWSEAHAVS GNGSRIVGAS GAGQGSVTAV IWESHLIKHL
     101  GTLGGEASSA EGISKDGEVV VGWSDTREGY THAFVFDGRD MKDLGTLGAT
     151  YSVARGVSGD GSIIVGVSAT ARGEDYGWQV GVKWEKGKIK QLKLLPQGLW
```

201 SEANAISEDG TVIVGRGEIS RNHIVAVKWN KNAVYSLGTL GGSVASAEAI
 251 SANGKVIWVW STTNNGETHA FMHKDETMHD LGTLGGGFVSV ATGVSADGRA
 301 IVGFSAVKTG EIHAFYYAEG EMEDLTTLGG BEARVFDISS BGNDIIGSIK
 351 TDAGAERAYL FHIHK*

- 5 A predicted signal peptide is highlighted.

The cp7108 nucleotide sequence <SEQ ID 82> is:

1 ATGAGTAAGA AGATAAAGGT TCTAGGTCAT TTGACGCTCT GCACTCTGTT
 51 TAGAGGAGTG CTGTGTGCAG CGGCCCTTTC CAACATAGGA TATGCGAGTA
 101 CTTCTCAGGA ATCACCATAT CAGAAGTCTA TAGAAGACTG GAAAGGGTAT
 151 ACCTTTACAG ATCTTGAGTT ACTGAGTAAG GAAGGGTGGT CTGAAGCTCA
 201 TGCAGTTTCT GGAATGGCA GTAGAATTGT AGGAGCTTCG GGAGCTGGCC
 251 AAGGTAGTGT GACTGCTGTC ATATGGGAAA GTCACCTGAT AAAACATCTC
 301 GGCACCTTTC GTGGCGAGGC TTCATCTGCA GAGGGAATTT CAAAGGATGG
 351 AGAGGTGGTC GTTGGGTGGT CAGATACTAG AGAGGGATAT ACTCATGCCT
 401 TTGTCTTCGA CGGTAGAGAT ATGAAAGATC TCGGTACTCT AGGAGCTACC
 451 TATTCTGTAG CAAGGGGTGT TTCTGGAGAT GGTAGTATCA TCGTAGGAGT
 501 CTCTGCAACT GCTCGTGGAG AGGATTACGG ATGGCAAGTT GGTGTCAAGT
 551 GGGAAAAAGG GAAATCAAA CAATTGAAGT TGTTCCTCA AGGTCTCTGG
 601 TCTGAGGCGA ATGCAATCTC TGAGGATGGT ACGGTGATTG TCGGGAGAGG
 651 GGAATCTCT CGCAATCACA TCGTTGCTGT AAAATGGAAT AAAATGCTG
 701 TGTATAGTTT GGGGACTCTC GGAGGTAGTG TCGCTTCAGC AGAGGCTATA
 751 TCGGCAATG GGAAGTAAT TGTAGGATGG TCCACGACTA ATAATGGTGA
 801 GACTCATGCC TTTATGCACA AAGATGAGAC AATGCACGAT CTCGGCACTC
 851 TAGGAGGAGG TTTTCTGTC GCAACTGGAG TTTCTGCTGA TGGGAGAGCC
 901 ATCGTAGGAT TTTCAGCAGT GAAGACCGGA GAAATTCATG CTTTTTACTA
 951 TGCAGAAGGA GAAATGGAGG ATTAAACAAC TTTGGGAGGG GAAGAAGCTC
 1001 GAGTGTTCGA CATATCTAGC GAAGGAAACG ATATCATTTG CTCTATAAAA
 1051 ACTGACGCTG GAGCTGAACG GCCTATCTG TTCCATATAC ATAAATAA

The PSORT algorithm predicts an outer membrane location (0.921).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 41A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 41B) and for FACS analysis (Figure 41C). A his-tagged protein was also expressed.

The cp7108 protein was also identified in the 2D-PAGE experiment.

- These experiments show that cp7108 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 42

The following *C.pneumoniae* protein (PID 4377287) was expressed <SEQ ID 83; cp7287>:

1 MVAKKTVRSY RSSFHSVIV AILSAGIAFE AHSLHSSELD LGVFNKQFEE
 51 HSAHVEEAQT SVLKGSDFVN PSQKESEKVL YTQVPLTQGS SGESLDLADA
 101 NFLEHFQHLF EETTVPFIDQ KLVWSDLDTR NFSQPTQEPD TSNAVSEKIS
 151 SDTKENRKDL ETEDPSKKSG LKEVSSDLPK SPETAVAAIS EDLEISENIS
 201 ARDPLQGLAF FYKNTSSQSI SEKSSSFQGI IFSGSGANS LGFENLKAPK
 251 SGAAVYSDRD IVFENLVKGL SFISCESLED GSAAGVNIV THCGDVTITD
 301 CATGLDLEAL RLVKDFSRGG AVFTARNHEV QNNLAGGILS VVGNGKAIVV
 351 EKNSAEKSNG GAFAGCSFVY SNNENTALWK ENQALSGGAI SSASDIDIQ
 401 NCSAIEFSGN QSLIALGEHI GLTDFVGGGA LAAQGTITLR NNAVVCVKV
 451 TSKTHGGAIL AGTVDLNETI SEVAFKQNTA ALTGGALSAN DKVIIANNFG
 501 EILFEQNEVR NHGGAICYGC RSNPKLEQKD SGENINIIGN SGATIFLKNK
 551 ASVLEVMTQA EDYAGGGALW GHNVLDSNS GNIQFIGNIG GSTFWIGEVV
 601 GGGAILSTDR VTISNNSGDV VFKNGKQCL AQKYVAPQET APVESDASST
 651 NKDEKSLNAC SHGDHYPPKT VEEVPPSL EEPVVSST IRGGGAILAQ
 701 HIFITDNTGN LRFSGNLGGG EESSTVGDLA IVGGGALLST NEVNVCNQ
 751 VVFSNDVTSN GCDSSGAILA KKVDISANHS VEFVSNNGSK FGGAVCALNE
 801 SVNITDNGSA VSPSKNRTRL GGAGVAAPQG SVTICGNQGN IAFKENFVFG

5 851 SENQRSGGGA IIANSSVNIQ DNAGDILFVS NSTGSYGGAI FVGSILVASEG
 901 SNPRTLITITG NSGDILFAKN STQTAASLSE KDSFSGGAIY TQNLKIVKNA
 951 GNVSFYGNRA PSGAGVQIAD GGTVCLEAFG GDILFEGNIN FDGSFNAIHL
 1001 CGNDSKIVEL SAVQDKNIIF QDAITYEENT IRGLPDKDVS PLSAPSLIFN
 1051 SKPQDDSAQH HEGTIRFSRG VSKIPQIAAI QEGTLALSQN AELWLAGLKQ
 1101 ETGSSIVLSA GSILRIFDSQ VDSSAPLPTE NKEETLVSAQ VQINMSSPTP
 1151 NKDKAVDTPV LADIISITVD LSSFVPEQDG TLPLPPEII PKGTKLHNSA
 1201 IDLKIIDPTN VGYENHALLS SHKDIPLISL KTAEGMTGTP TADASLSNIK
 1251 IDVSLPSITP ATYGHTGVWS ESKMEDGRLV VGWQPTGYKL NPEKQCALVL
 15 1301>NNLWSHYTDL RALKQEIFAH HTIAQRMELD FSTNVWGSGL GVVEDCQNIQ
 1351>EFDGFKHHLT GYALGLDTQL VEDFLIGGCF SQFFGKTESQ SYKAKNDVKS
 1401>YMGAAYAGIL AGPWLIKAF VYGNINNDLT TDYGLTGIST GSWIGKGFIA
 1451>GTSIDYRYIV NRRRFISAIV STVVPFVEAE YVRIDLPEIS EQKEQVRLFQ
 1501>KTRFENVAIK FGFALHAYS RGSRAEVNSV QLAVVFDVYR KGPVSLITLK
 1551>DAAYSWKSYG VDIPLKAWKA RLSNNTWNS YLSTYLAFNY EWREDLIAYD
 1601>FNGGIRIIF*

A predicted signal peptide is highlighted.

The cp7287 nucleotide sequence <SEQ ID 84> is:

20 1 ATGGTAGCGA AAAAAACAGT ACGATCTTAT AGGTCTTCAT TTTCTCATTC
 51 CGTAATAGTA GCAATATTGT CAGCAGGCAT TGCTTTTGAA GCACATTCCT
 101 TACACAGCTC AGAAGTAGAT TTAGGTGTAT TCAATAAACA GTTTGAGGAA
 151 CATTCTGCTC ATGTTGAAGA GGCTCAAACA TCTGTTTAA AGGGATCAGA
 201 TCCTGTAAAT CCCTCTCAGA AAGAATCCGA GAAGGTTTTG TACACTCAAG
 251 TGCCCTCTAC CCAAGGAAGC TCTGGAGAGA GTTTGGATCT CGCCGATGCT
 301 AATTCTTAG AGCATTTCAT GCATCTTTT GAAGAGACTA CAGTATTTGG
 351 TATCGATCAA AAGCTGGTTT GGTCAGATT AGATACTAGG AATTTTCCC
 401 AACCCTCTCA AGAAGCTGAT ACAAGTAATG CTGTAAGTGA GAAAATCTCC
 451 TCAGATACCA AAGAGAATAG AAAAGACCTA GAGACTGAAG ATCCTTCAAA
 501 AAAAAGTGGC CTTAAAGAAG TTTCATCAGA TCTCCCTAAA AGTCTGAAA
 30 551 CTGCAGTAGC AGCTATTTCT GAAGATCTTG AAATCTCAGA AAACATTTCA
 601 GCAAGAGATC CTCTTCAGGG TTTCAGCATTT TTTTATAAAA ATACATCTTC
 651 TCACTCTATC TCTGAAAAGG ATTCTTCATT TCAAGGAATT ATCTTTCTG
 701 GTTCAGGAGC TAATTCAGGG CTAGGTTTTG AAAATCTTAA GGCGCCGAAA
 35 751 TCTGGGGCTG CAGTTTATTC TGATCGAGAT ATTGTTTTTG AAAATCTTGT
 801 TAAAGGATG AGTTTTATAT CTGTGGAATC TTAGAAGAT GGCTCTGCCG
 851 CAGGTGTAAG CATTGTTGTG ACCCATTTGT GTGATGTAAC TCTCACTGAT
 901 TGTGCCACTG GTTTAGACCT TGAAGCTTTA CGTCTGTTA AAGATTTTTC
 951 TCGTGGAGGA GCTGTTTTCA CTGCTCGCAA CCATGAAGTG CAAAATAACC
 1001 TTGCAGGTGG AATTCTATCC GTTGTAGGCA ATAAAGGAGC TATTGTTGTA
 40 1051 GAGAAAAATA GTGCTGAGAA GTCCAATGGA GGAGCTTTTG CTGCGGAAG
 1101 TTTTGTATAC AGTAACAACG AAAACACCGC CTGTGGAAGG GAAAATCAAG
 1151 CATATATCAG AGGAGCCATA TCCTCAGCAA GTGATATTGA TATTCAAGGG
 1201 AACTGTAGCG CTATTGAATT TTCAGGAAAC CAGTCTCTAA TTGCTCTTGG
 1251 AGAGCATATA GGGCTTACAG ATTTTGTAGG TGGAGGAGCT TTAGCTGCTC
 45 1301 AAGGGACGCT TACCTTAAGA AATAATGCAG TAGTGCAATG TGTAAAAAC
 1351 ACTTCTAAAA CACATGGTGG AGCTATTTTA GCAGGTACTG TTGATCTCAA
 1401 CGAAACAATT AGCGAAGTTG CCTTTAAGCA GAATACAGCA GCTCTAACTG
 1451 GAGGTGCTTT AAGTGCAAAAT GATAAGGTTA TAATTGCAAA TAACTTTGGA
 50 1501 GAAATTCTTT TTGAGCAAAA CGAAGTGAGG AATCACGGAG GAGCCATTTA
 1551 TTGTGGATGT CGATCTAATC CTAAGTTAGA ACAAAGGAT TCTGGAGAGA
 1601 ACATCAATAT TATTGGAAAC TCCGGAGCTA TCACTTTTAA AAAAAATAAG
 1651 GCTTCTGTTT TAGAAGTGAT GACACAAGCT GAAGATTATG CTGGTGGAGG
 1701 CGCTTTATGG GGGCATAATG TTCTTCTAGA TTCCAATAGT GGAATATTC
 55 1751 AATTTATAGG AATATAGGT GGAAGTACCT TCTGGATAGG AGAATATGTC
 1801 GGTGGTGGTG CGATTCTCTC TACTGATAGA GTGACAATTT CTAATAACTC
 1851 TGGAGATGTT GTTTTAAAG GAAACAAAGG CCAATGTCTT GCTCAAAAAAT
 1901 ATGTAGCTCC TCAAGAAACA GCTCCCGTGG AATCAGATGC TTCATCTACA
 1951 AATAAGACG AGAAGAGCCT TAATGCTTGT AGTCATGGAG ATCATTATCC
 2001 TCCTAAAACT GTAGAAGAGG AAGTGCCACC TTCATTGTGA GAAGAACATC
 60 2051 CTGTTGTTTC TTCGACAGAT ATTCGTGGTG GTGGGGCCAT TCTAGCTCAA
 2101 CATATCTTTA TTACAGATAA TACAGGAAAT CTGAGATTCT CTGGGAACCT
 2151 TGGTGGTGGT GAAGAGTCTT CTAATGTCGG TGATTTAGCT ATCGTAGGAG
 2201 GAGGTGCTTT GCTTTCTACT AATGAAGTTA ATGTTTGCAG TAACCAAAAT
 2251 GTTGTTTTTC CTGATAACGT GACTTCAAAT GGTGTGATT CAGGGGGAGC
 65 2301 TATTTTAGCT AAAAAAGTAG ATATCTCCGC GAACCACTCG GTTGAATTTG

2351 TCTCTAATGG TTCAGGGAAA TTCGGTGGTG CCGTTTGCGC TTAAACGAA
 2401 TCAGTAAACA TTACGGACAA TGGCTCGGCA GTATCATCTT CTAAAAATAG
 2451 AACACGTCCTT GCGGGTGCCT GAGTTGCAGC TCCTCAAGGC TCTGTAACGA
 2501 TTTGTGGAAA TCAGGGAAAC ATAGCATTTA AAGAGAAGCT TGTTTTGGC
 5 2551 TCTGAAAATC AAAGATCAGG TGGAGGAGCT ATCATTGCTA ACTCTTCTGT
 2601 AAATATTTCAG GATAACGCAG GAGATATCCT ATTTGTAAGT AACTCTACGG
 2651 GATCTTATGG AGGTGCTATT TTTGTAGGAT CTTTGGTTGC TTCTGAAGGC
 2701 AGCAACCCAC GAACGCTTAC AATTACAGGC AACAGTGGGG ATATCCTATT
 10 2751 TGCTAAAAAT AGCACGCAAA CAGCCGCTTC TTTATCAGAA AAAGATTCCCT
 2801 TTGGTGGAGG GGCATCTAT ACACAAAACC TCAAAATTGT AAAGAATGCA
 2851 GGAACGTTT CTTTCTATGG CAACAGAGCT CTTAGTGGTG CTGGTGTCCA
 2901 AATTGCAGAC GGAGGAACGT TTTGTTTAGA GGCCTTTTGA GGAGATATCT
 2951 TATTTGAAGG GAATATCAAT TTTGATGGGA GTTCAATGC GATTCACTTA
 15 3001 TGCGGGAATG ACTCAAAAAT CGTAGAGCTT TCTGCTGTTC AAGATAAAAA
 3051 TATTTATTTT CAAGATGCAA TTACTTATGA AGAGAACACA ATTCTGGGCT
 3101 TGCCAGATAA AGATGTCACT CTTTAAAGTG CCCCTTCATT AATTTTAAAC
 3151 TCCAAGCCAC AAGATGACAG CGCTCAACAT CATGAAGGGA CGATACGGTT
 3201 TTCTCGAGGG GTATCTAAAA TTCCCTCAGT TGCTGCTATA CAAGAGGGAA
 20 3251 CCTTAGCTTT ATCACAAGAC GCAGAGCTTT GGTGGCAGG ACTTAAACAG
 3301 GAAACAGGAA GTTCTATCGT ATTGTCTGCG GGATCTATTC TCCGTATTTT
 3351 TGATTCCAG GTTGATAGCA GTGCGCCTCT TCCTACAGAA AATAAAGAGG
 3401 AGACTCTTGT TTCTGCCGGA GTTCAAATTA ACATGAGCTC TCCTACACCC
 3451 AATAAAGATA AAGCTGTAGA TACTCCAGTA CTTGCAGATA TCATAAGTAT
 25 3501 TACTGTAGAT TTGTCTTCAT TTGTTCTTGA GCAAGACGGA ACTCTTCTCT
 3551 TTCTCTCTGA AATTATCATT CCTAAGGGAA CAAAATTACA TTCTAATGCC
 3601 ATAGATCTTA AGATTATAGA TCCTACCAAT GTGGGATATG AAAATCATGC
 3651 TCTTCTAAGT TCTCATAAAG ATATTCCATT AATTTCTCTT AAGACAGCGG
 3701 AAGGAATGAC AGGGACGCTT ACAGCAGATG CTTCTCTATC TAATATAAAA
 3751 ATAGATGTAT CTTTACCTTC GATCACACCA GCAACGTATG GTCACACAGG
 30 3801 AGTTTGGTCT GAAAGTAAAA TGGAAGATGG AAGACTTGTA GTCGTTGGC
 3851 AACCTACGGG ATATAAGTTA AATCCTGAGA AGCAAGGGGC TCTAGTTTGT
 3901 AATAATCTCT GGAGTCATTA TACAGATCTT AGAGCTCTTA AGCAGGAGAT
 3951 CTTTGCTCAT CATACGATAG CTCAAAGAA GTGAGTTAGAT TTCTCGACAA
 35 4001 ATGTCTGGGG ATCAGGATTA GGTGTTGTTG AAGATTGTCA GAACATCGGA
 4051 GAGTTTGATG GGTTCAAACA TCATCTCACA GGGTATGCCC TAGGCTTGGA
 4101 TACACAACCTA GTTGAAGACT TCTTAATTGG AGGATGTTTC TCACAGTTCT
 4151 TTGGTAAAAA TGAAAGCCAA TCCTACAAG CTAAGAACGA TGTGAAGAGT
 4201 TATATGGGAG CTGCTTATGC GGGGATTTTA GCAGGTCCTT GGTAAATAAA
 4251 AGGAGCTTTT GTTTACGGTA ATATAAACAA CGATTGACT ACAGATTACG
 40 4301 GTACTTTAGG TATTTCAACA GGTTCATGGA TAGGAAAAGG GTTTATCGCA
 4351 GGCACAAGCA TTGATTACCG CTATATTGTA AATCCTCGAC GGTTTATATC
 4401 GGCAATCGTA TCCACAGTGG TTCCTTTTGT AGAAGCCGAG TATGTCCGTA
 4451 TAGATCTTCC AGAAATTAGC GAACAGGGTA AAGAGGTTAG AACGTTCCAA
 4501 AAAACTCGTT TTGAGAAATG CGCCATTCCCT TTTGGATTTG CTTTAGAACA
 45 4551 TGCTTATTCG CGTGGCTCAC GTGCTGAAGT GAACAGTGTA CAGCTTGCTT
 4601 ACGTCTTTGA TGTATATCGT AAGGGACCTG TCTCTTTGAT TACACTCAAG
 4651 GATGCTGCTT ATTCTTGGA GAGTTATGGG GTAGATATTC CTTGTAAGGC
 4701 TTGGAAGGCT CGCTTGAGCA ATAATACGGA ATGGAATTCA TATTTAAGTA
 4751 CGTATTTAGC GTTTAATTAT GAATGGAGAG AAGATCTGAT AGCTTATGAC
 50 4801 TTCAATGGTG GTATCCGTAT TATTTTCTAG

The PSORT algorithm predicts an inner membrane location (0.106).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 42A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 42B) and for FACS analysis (Figure 42C). A his-tagged protein was also expressed.

55 The cp7287 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7287 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 43

The following *C.pneumoniae* protein (PID 4377105) was expressed <SEQ ID 85; cp7105>:

```

1  MSLYQKWNNS QLKKSLEYST VAALIFMIPS QESFADSLID LNLGLDPSVE
51  CLSGDGAFSV GYFTKAGSTP VEYQPFKYDV SKKTFITLSV ETANQSGYAY
101 GISYDGTITV GTCSLGAGKY NGAKWSADGT LTPLTGITGG TSHTEARAIS
151 KDTQVIEGFS YDASGQPKAV QWASGATTVT QLADISGGSR SSYAYAISDD
201 GTIIVGSMES TITRKTAVK WVNVPYTLG TLGGDASTGL YISGDGTIVV
251 GAANTATVTN GNQESHAYMY KDNQMKD*

```

The cp7105 nucleotide sequence <SEQ ID 86> is:

```

10 1  GTGAGTCTAT ATCAAAAATG GTGGAACAGT CAGTTAAAGA AGAGCCTCTG
51  CTATTCGACT GTTGCTGCTC TAATATTTAT GATTCCTTCT CAAGAATCCT
101 TTGCAGATAG TCTTATAGAT TTAAATTTAG GTTTAGATCC TTCGGTCGAA
151 TGTCTGTCAG GAGATGCTGC ATTTTCTGTT GGGTATTTTA CTAAGGCGGG
15 201 ATCGACTCCC GTAGAATATC AGCCGTTTAA ATACGACGTA TCTAAGAAGA
251 CATTACAAT CCTTTCCGTA GAAACGGCAA ATCAGAGCGG CTATGCTTAC
301 GGAATCTCCT ACGATGGCAC GATCACTGTA GGAACGTGTA GCCTAGGTGC
351 AGGAAAATAT AACGGCGCAA AATGGAGTGC GGATGGCACT TTAACACCCCT
401 TAACTGGAAT CACGGGGGGG ACGTCACATA CGGAAGCGCG TGCGATTTCCT
451 AAGGATACTC AGGTGATCGA GGGTTTCTCA TATGATGCTT CAGGGCAACC
20 501 CAAGGCTGTG CAGTGGGCAA GCGGAGCGAC TACAGTAACA CAATTAGCAG
551 ATATTTTCAGG AGGCTCTAGA AGCTCTTATG CGTATGCTAT ATCTGATGAT
601 GGCACGATTA TTGTTGGGTC TATGGAGAGC ACGATAACAA GGAAAACACTAC
651 AGCTGTAAAA TGGGTAAATA ATGTTCTTAC GTATCTGGGA ACCTTAGGAG
701 GAGATGCTTC TACAGTCTT TATATTTCTG GAGACGGCAC CGTGATGTGA
25 751 GGTGCGCAA ATACAGCAAC TGTAACCAAT GGAATCAGG AATCCCACGC
801 CTATATGTAT AAAGATAACC AAATGAAAGA TTGA

```

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 43A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 43B) and for FACS analysis (Figure 43C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7105 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 44

The following *C.pneumoniae* protein (PID 4376802) was expressed <SEQ ID 87; cp6802>:

```

1  MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLLIEGK
51  LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
101 ATLESRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLRQTP ENYDGLLLIG
40 151 DAALQHPVLP GFVTYDLASG WYDLTKLPFV FALLLHSTW KEHPLPNLAM
201 EEALQQFESS PEEVLKEAHQ HTGLPPSLQ EYVALCQYRL GEEHYESFEK
251 FREYYGTLYQ QARL*

```

A predicted signal peptide is highlighted.

The cp6802 nucleotide sequence <SEQ ID 88> is:

```

45 1  ATGTCTAACC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
51  TAATTCCTTT CCGCTGTCCC TACAACTCAT AAAAAGAAAC GATATTCGCT
101 GTGTTCTTGC TCCCCCTGCA GACCTCCTCA ACTTGCTAAT CGAAGGGAAA
151 CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTGGG
201 GTATGTCCCC GGCTTTGGAA TTGCAGCAA CCAACGTATC CTCAGTGTAA

```


5
10
20
25
30
35
40
45
50
55
60
65
70
75

```

251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC
301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAAG TGCTTTGTCTG
351 TCATCTCTGG CGCATCCCAA CTCCTCATAT CCTAAGATTG ATAACACAA
401 AAGTACTCAG ACAAACCCCT GAAAATTATG ATGGCCTCCT CCTAATCGGA
451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTGTGTA CCTATGACCT
501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTTGCTCTTC
551 TTCTACACAG CACCTCTTGG AAAGAACATC CCTACCCAA CCTTGCGATG
601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA
651 AGCTCATCAA CATACAGGTC TGCCCCCTTC TCTTCTTCAA GAATACTATG
701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAA
751 TTCCGGGAAT ATTATGGAAC CCTCTACCAA CAAGCCCGAC TGTAA

```

The PSORT algorithm predicts an inner membrane location (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 44A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 44B) and for FACS analysis (Figure 44C). A his-tagged protein was also expressed.

These experiments show that cp6802 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 45

The following *C.pneumoniae* protein (PID 4376390) was expressed <SEQ ID 89; cp6390>:

20
25
30
35

```

1 MVFSYYCMGL FFFSGAISSC GLLVSLGVGL GLSVLGVLLL LLAGLLLLFKI
51 QSMLEVPKFA PDLLDLEDAS ERLRVKASRS LASLPKEISQ LESYIRSAAN
101 DLNTIKTWPH KDQRLVETVS RKLERLAAQ NYMISELCEI SEILEEEHH
151 LILAQESLEW IGKSLFSTFL DMESPLNLSH LSEVRPYLAV NDPRLLEITE
201 ESWEVVSHEI NVTSAFKKAQ ILFKNNEHSR MKKKLESVQE LLETFTYKSL
251 KRSYRELGLCL SEKMRIHDN PLFPWVQDQQ KYAHAKNEFG EIAKCLEEPE
301 KTFFWLDEEC AISYMDCWDF LNESIQNKKS RVDYDYISTK KIALKDRART
351 YAKVLLLENP TTEGKIDLQD AQRATERQSQ EFTLEHTET KVRLEALQQC
401 FSDLREATNV RQVRFTNSEN ANDLKESEK IDKERVRYQK EQRLYWETID
451 RNEQELREBI GESLRLQNRK KGYRAGYDAG RLKGLLRQWK KNLRDVEAHL
501 EDATMDFEHE VSKSELCSVR ARLEVLEBEL MDMSPKVADI EELLSYBERC
551 ILPIRENLER AYLYQNKCE ILSKAKFFFP EDEQLLVSEA NLREVGAQLK
601 QVQKCKQERA QKFAIFEKHI QEQLSLIKEQ VRSFDLAGVG FLKSELLSIA
651 CNLYIKAVVK ESTPVDVPCM QLYSYIEDN EAVVRNRLN MTERYQNFKR
701 SLNSIQFNGD VLLRDPVYQP EGHETRLKER ELQETTLCK KLVQAQDRLS
751 ELESRLSRR

```

A predicted signal peptide is highlighted.

The cp6390 nucleotide sequence <SEQ ID 90> is:

40
45
50
55

```

1 TTGGTATTCT CATACTATTG CATGGGATTA TTTTCTTCT CTGGAGCTAT
51 TTCTAGTTGT GGTCTTTTAG TGTCTCTAGG AGTTGGTTTA GGACTTAGTG
101 TTTTAGGAGT ACTTTTACTT CTCTTAGCAG GTCTTTTGCT TTTTAAGATC
151 CAAAGTATGC TTCGAGAGGT GCCTAAGGCT CCTGATCTAT TAGATTAGA
201 AGATGCAAGT GAACGGCTTA GAGTAAAGGC TAGCCGTCTT TTAGCAAGCC
251 TCCCGAAGGA AATCAGTCAG CTAGAGAGCT ACATTCGTTC TGCAGCTAAT
301 GATCTAAATA CAATTAAGAC TTGGCCGCAT AAAGATCAAA GACTCGTCGA
351 GACCGTGTCA CGAAAATTAG AGCGTCTGGC AGCTGCTCAA AACTATATGA
401 TTTCTGAACT CTGCGAGATT AGTGAGATTC TTGAGGAAGA GGAGCATCAT
451 CTAATTTTGG CTCAGGAATC TCTAGAATGG ATAGGTAAGA GTCTATTTTC
501 TACCTTTCTG GACATGGAAT CTTTTTTAAA TTTGAGCCAT CTATCTGAAG
551 TGCGTCCGTA CTTAGCTGTA AATGATCCTA GATTATTAGA AATTACCGAA
601 GAATCTTGGG AAGTAGTGAG TCATTTTATA AATGTAACGT CTGCTTTTAA
651 GAAAGCTCAG ATTCTTTTAA AGAACACGA ACATTTCTCGG ATGAAGAAGA
701 AGTTAGAAAG TGTTCAGAG TTAAGTGAAG CATTTATTTA TAAGAGTTTA
751 AAGAGAAGTT ATCGAGAATT AGGATGCTTA AGTGAAGAAG TGAGAATCAT
801 TCACGACAAT CCTCTTTTCC CTTGGGTGCA AGATCAGCAG AAGTATGCTC
851 ATGCTAAGAA TGAATTTGGA GAGATTGCGC GGTGTTTAGA GGAGTTTGAA
901 AAGACGTTCT TCTGGTTGGA TGAGGAGTGT GCTATTTCTT ACATGGACTG

```


5 951 TTGGGATTTT CTAAATGAGT CTATTCAGAA TAAGAAGTCC AGAGTAGATC
 1001 GAGATTATAT ATCCACGAAG AAAATTGCAT TAAAGGATAG AGCCCCGCACT
 1051 TATGCTAAGG TTCTTTTAGA AGAGAATCCG ACTACAGAGG GTAAAATAGA
 1101 TTGCAAGAC GCTCAAAGAG CCTTTGAGCG TCAAAGTCAG GAGTTTATATA
 1151 CACTAGAGCA TACGGAACA AAGGTGAGAC TAGAAGCACT TCAACAGTGC
 1201 TTCTCGGATC TTAGGGAGGC GACGAACGTA AGGCAAGTTA GGTTTACAAA
 1251 TTCTGAAAAT GCGAATGATT TAAAGGAGAG TTTCGAGAAG ATAGATAAAG
 1301 AGCGTGTGCG ATATCAAAAA GAGCAAAGGC TCTATTGGGA AACAATAGAT
 1351 CGCAATGAGC AAGAGCTTAG GGAAGAGATT GGGGAGTCGC TTCGTTTACA
 1401 AAATCGGAGA AAAGGGTATA GGGCTGGATA TGATGCTGGG CGTTTAAAG
 1451 GTTGTGTCG TCAGTGGAG AAAAATCTCC GCGATGTGGA AGCCCACCTT
 1501 GAAGATGCAA CTATGGATT TTGAGCATGAA GTAAGCAAGA CGGAATTGTG
 1551 CAGTGTTCGG GCGAGGCTCG AGGTTCCTAGA AGAAGAGCTG ATGGATATGT
 1601 CTCCTAAAGT TCGGATATA GAAGAGTTGT TGTCCATGA AGAGCGTTGT
 1651 ATTCTTCCTA TTAGGAAAA TTTAGAAAGG GCATACCTCC AATATAATAA
 1701 GTGTTCTGAA ATTTTATCCA AGGCAAAGTT CTCTTTCCG GAAGCAGCAG
 1751 AATTGCTAGT TTCGGAAGCG AATCTAAGAG AGGTGGGTGC CCAGTTAAAA
 1801 CAAGTACAGG GAAAAATGCA AGAGAGGGCC CAAAAGTTCG CAATATTTGA
 1851 AAAGCATAT CAGGAGCAGA AAAGCCTTAT TAAAGAGCAA GTGCGGAGTT
 1901 TTGATCTAGC GGGAGTTGGG TTTTAAAGA GTGAGCTTCT TAGTATGTCT
 1951 TGTAACCTTT ATATAAGGC GGTGTTAAG GAGTCTATAC CAGTTGATGT
 2001 GCCTTGATG CAGTTATATT ATAGTTATTA CGAAGATAAT GAAGCTGTAG
 2051 TCGGAAACCG CCTTTTAAAT ATGACGGAGA GGTATCAAAA TTTTAAAGG
 2101 AGTTTGAATT CCATACAATT TAATGGTGAC GTTCTTTTAC GGGATCCGGT
 2151 CTATCAACCT GAAGGTCATG AGACCAGGCT AAAGGAACGG GAGCTACAAG
 2201 AAACAACCTT GTCTTGTAAG AAATTAAAAG TGGCTCAAGA TCGTCTTCT
 2251 GAATTAGAGT CAAGGCTGTC TAGGAGATAG

The PSORT algorithm predicts a periplasmic location (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 45A.

30 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 45B) and for FACS analysis (Figure 45C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

35 These experiments show that cp6390 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 46

The following *C.pneumoniae* protein (PID 4376272) was expressed <SEQ ID 91; cp6272>:

1 MKRCFLFLAS FVLMGSSADA LTHQEAVKKK NSYLSHFVSV SGIVTIEDGV
 51 LNIHNNLRIQ ANKVYVENTV GQSLKLVAHG NVMVNYRAKT LVC DYLEYE
 40 101 DTDSCLLTNG RFAMPWFLG GSMITLTPET IVIRKGYIST SEGPKDLCCL
 151 SGDYLEYSSD SLLSIGKTTL RVCRIPIFL PPSIMPMEI PKPPINFRGG
 201 TGGFLGSLG MSYSPISRKH FSSTFFLDSF FKHGVMGMFN LHCSQKQVPE
 251 NVFNMKSYA HRLAIDMAEA HDYRLHGDF CFTHKHVNFV GEYHLSDSWE
 301 TVADIFPNF MLKNTGPTRV DCTWNDNYFE GYLTSVSVKN SFQANQELP
 45 351 YLTLRQYPIS IYNTGVYLEN IVECGYLNFA FSDHIVGENF SSLRLAARPK
 401 LHKTVPPLIG TSSSTLGSSL IYSDVPEIS SRHSQLSAKL QLDYRFLHKK
 451 SYIQRRIIE PFVTFITETR PLAKNEDHYI FSIQDAFHSI NLLKAGIDTS
 501 VLSKTNPRFP RIHAKLWTH ILSNTESKPT FPKTACELSL PFGKNTVSL
 551 DAEWIWKKHC WDHMNIWREW IGNDNVAMTL ESLHRSKYSI IKCDRENFIL
 50 601 DVSRPIDQLL DSPLSDHRNL ILGKLFVRPH PCWNYRLSLR YGWHRQDTPN
 651 YLEYQMILGT KIFEHWQLYG VYERREADSR FFFFLKLDKP KKPPF*

A predicted signal peptide is highlighted.

The cp6272 nucleotide sequence <SEQ ID 92> is:

1 ATGAAACGTT GCTTCTTATT TCTAGCTTCC TTTGTTCTTA TGGGTTCCCTC

51 AGCTGATGCT TTGACTCATC AAGAGGCTGT GAAAAAGAAA AACTCCTATC
 101 TTAGTCACTT TAAGAGTGT TCTGGGATG TGACCATCGA AGATGGGGTA
 151 TTGAATATCC ATAACAACCT GCGGATACAA GCCAATAAAG TGTATGTAGA
 201 AAATACTGTG GGTCAAAAGCC TGAAGCTTGT CGCACATGGC AATGTTATGG
 251 TGAACATATAG GGCAAAAACC CTAGTTTGTG ATTACCTAGA GTATTACGAA
 301 GATACAGACT CTTGTCTTCT TACTAATGGA AGATTTCGCGA TGTATCCTTG
 351 GTTCTTAGGG GGGTCTATGA TCACTCTAAC CCCAGAAAACC ATAGTCATTC
 401 GGAAGGGATA TATCTCTACC TCCGAGGGTC CCAAAAAGA CCTGTGCCTC
 451 TCCGGAGATT ACCTGGAATA TTCTTCAGAT AGTCTTCTTT CTATAGGGAA
 501 GACAACATTA AGGGTGTGTC GCATTCCGAT ACTTTTCTTA CCTCCATTTT
 551 CTATCATGCC TATGGAGATC CCTAAGCCTC CGATAAACTT TCGAGGAGGA
 601 ACAGGAGGAT TTCTGGGATC CTATTTGGGG ATGAGCTACT CGCCGATTTC
 651 TAGGAAGCAT TTCTCCTCGA CATTTTCTTT GGATAGCTTT TTCAAGCATG
 701 GCGTCGGCAT GGGATTCAAC CTCCATTGTT CTCAGAAGCA GGTTCCTGAG
 751 AATGCTTCA ATATGAAAAG CTATTATGCC CACCGCCTTG CTATCGATAT
 801 GGCAGAAGCT CATGATCGCT ATCGCCTACA CGGAGATTTC TGCTTCACGC
 851 ATAAGCATGT AAATTTTCTT GGAGAATACC ATCTCAGCGA TAGTTGGGAA
 901 ACTGTTGCTG ACATTTTCCC CAACAACCTT ATGTTGAAAA ATACAGGCCC
 951 CACACGTGTC GATTGCACTT GGAATGACAA CTATTTTGAA GGGTATCTCA
 1001 CCTCTTCTGT TAAGGTAAAC TCTTTCCAAA ATGCCAACCA AGAGCTCCCT
 1051 TATTTAACAT TAAGGCAGTA CCCGATTCTT ATTTATAATA CGGGAGTGTA
 1101 CCTTGAAAAC ATCGTAGAAT GTGGGTATTT AAACCTTGCT TTTAGCGATC
 1151 ATATCGTTGG CGAGAATTTT TCTTCACTAC GTCTTGCTGC GCGCCCTAAG
 1201 CTCCATAAAA CTGTGCTCTT ACCTATAGGA ACGCTCTCCT CCACCTAGG
 1251 GAGTTCTCTG ATTTACTATA GCGATGTTC TGAGATCTCC TCGCGCCATA
 1301 GTCAGCTTTC CGCGAAGCTA CAACTTGATT ATCGCTTCT ATTACATAAG
 1351 TCCTACATTC AAAGACGCCA TATTATAGAG CCGTTCGTTA CCTTCATTAC
 1401 AGAGACTCGT CCTCTAGCTA AGAATGAAGA TCATTATATC TTTTCTATTTC
 1451 AAGATGCCTT TCACTCCTTA AACCTTCTGA AAGCGGGTAT AGATACCTCG
 1501 GTACTGAGTA AGACTAACCC TCGATTCCCG AGAATCCATG CGAAGCTGTG
 1551 GACTACCCAC ATCTTGAGCA ATACAGAAAG CAAACCCACG TTTCCCAAAA
 1601 CTGCATGCGA GCTATCTCTA CCTTTTGGA AGAAAAATAC AGTCTCCTTA
 1651 GATGCTGAAT GGATTTGGA AAAGCACTGT TGGGATCACA TGAACATACG
 1701 TTGGGAGTGG ATCGGAAATG ACAAATGTGG TATGACTCTA GAATCCCTGC
 1751 ATAGAAGCAA ATACAGCCTG ATTAAGTGTG ACAGGGAGAA CTTTATTTTA
 1801 GATGTCAGCC GTCCCATTTGA CCAGCTTTTA GACTCCCTC TCTCTGATCA
 1851 TAGGAATCTC ATTTTAGGGA AATTATTTGT ACGACCTCAT CCTGTGTTGA
 1901 ATTACCGCTT ATCCTTACGC TATGGCTGGC ATCGCCAGGA CACTCCGAAC
 1951 TACCTAGAAT ACCAGATGAT TCTAGGGACG AAGATCTTCG AACATTGGCA
 2001 GCTCTATGGG GTGTATGAAC GCCGAGAAGC AGATAGTCTGA TTTTCTTCTT
 2051 TCTTAAAGCT CGACAAACCT AAAAAACCTC CCTTCTAA

The PSORT algorithm predicts an outer membrane location (0.48).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 46A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 46B). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6272 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

50 Example 47

The following *C.pneumoniae* protein (PID 4377111) was expressed <SEQ ID 93; cp7111>:

1 MFEAVIADIQ AREILDSRGY PTLHVKVTTT TGSVGEARVP SGASTGKKEA
 51 LEFRDTSQPR YQKGVLQAV KNVKELFPL VKGCSVYEQS LIDSLMMDSD
 101 GSPNKETLGA NAILGVSLAT AHAAATLRR PLYRYLGGCF ACSLPCPMMN
 151 LINGMHADN GLEFQEFMIR PIGASSIKEA VMGADVFTT LKKLLHERGL
 201 STGVGDEGGF APNLASNEEA LELLLLAIEK AGFTPGKDIS LALDCAASSF

251 YNVKTGTYDG RHYEEQIAIL SNLCDRYPID SIEDGLAEED YDGWALLTEV
 301 LGEKVQIVGD DLFVTNPPELI LEGISNGLAN SVLIKPNQIG TLTETVYAIK
 351 LAQMAGYTTI ISHRSGETTD TTIADLAVAF NAGQIKTGSL SRSERVAKYN
 401 RLMEIEEELG SEAIPTDSNV FSYEDSEE*

5 A predicted signal peptide is highlighted.

The cp7111 nucleotide sequence <SEQ ID 94> is:

1 ATGTTTGAAG CTGTCATTGC CGATATCCAG GCTAGGGAAA TCTTGGATTCT
 51 TCGCGGGTAT CCCACTTTAC ATGTTAAAGT AACCAC TAGC ACAGGTTCTG
 101 TTGGAGAAGC TCGGGTTCCCT TCAGGAGCAT CCACAGGGAA AAAAGAAGCC
 151 TTAGAGTTTC GTGATACAGA TTCTCCTCGT TATCAAGGCA AAGGGGTTT
 201 GCAAGCTGTA AAAACGTAA AAGAAATCT TTTCCCTC GTCAAGGGAT
 251 GTAGTGTFTA TGAGCAATCC TTAATTGATT CTCTGATGAT GGATTCTGAC
 301 GGCTCTCCGA ACAAAGAAAC TCTAGGGGCC AATGCTATTT TAGGAGTCTC
 351 TCTAGCTACA GCACATGCAG CAGCAGCAAC ACTACGCAGA CCTCTGTATC
 401 GTTATTTAGG AGGGTGTTTT GCCTGCAGTC TTCCCTGTCC TATGATGAAT
 451 CTGATCAATG GAGGCATGCA TGCCGATAAC GGCTTGGAGT TCCAAGAATT
 501 TATGATCCGT CCTATTGGAG CCTCTTCCAT CAAAGAAGCT GTCAACATGG
 551 GTGCTGACGT TTTTCATACT TTGAAAAAAT TACTCCATGA AAGAGGCTTA
 601 TCTACTGGAG TGGGTGACGA AGGAGGCTTC GCCCGAATC TTGCTTCTAA
 651 TGAAGAAGCT CTAGAGCTCC TATTGCTGGC TATTGAAAAA GCAGGCTTTA
 701 CTCCAGGAAA AGATATATCG CTAGCCCTAG ACTGCGCAGC ATCCTCATTC
 751 TATAACGTAA AAACAGGCAC GTATGATGGG AGGCACTATG AAGAGCAAAT
 801 CGCAATCCTT TCTAATTTAT GTGATCGCTA TCCTATAGAC TCCATAGAAG
 851 ATGGTCTTGC TGAAGAAGAC TATGACGGGT GGGCCTTGTT AACTGAAGTT
 901 CTGGAGAAA AAGTACAGAT TGTGGGTGAT GACCTATTG TTACAAATCC
 951 GGAATTAATA TTAGAGGGTA TTAGCAATGG ATTAGCGAAC TCTGTGTTGA
 1001 TTAAACCAA TCAGATAGGG ACGCTTACTG AAACAGTGTA TGCTATCAAG
 1051 CTTGCGCAA TGGCTGGCTA TACTACAATT ATTTCTCATC GCTCAGGAGA
 1101 AACTACGGAC ACTACGATTG CAGATCTTGC TGTGCTTTC AACGCCGGTC
 1151 AAATCAAAAC AGGCTCTTTA TCACGTTCTG AGCGTGTGTC AAAATACAAT
 1201 AGACTCATGG AAATGAAGA AGAGCTTGA TCCGAAGCAA TTTTCACAGA
 1251 TTCTAATGTA TTTTCTTAC GAGGATTCT GAGGAATAG

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 47A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 47B) and for FACS analysis (Figure 47C). A his-tagged protein was also expressed.

The cp7111 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

40 These experiments show that cp7111 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 48

The following *C.pneumoniae* protein (PID 4455886) was expressed <SEQ ID 95; cp0010>:

1 MKSQFSWLVL SSTLACFTSC STVFAATAEN IGPDSDFDGS TNTGTYTPKN
 45 51 TTTGIDYTLT GDITLQNLGD SAALTKGCFS DTPESLSFAG KGYSLSFLNI
 101 KSSAEGAALS VTDDKNSLT GFSSLTFLAA PSSVITTPSG KGAVKCGGDL
 151 TFDNNGTILF KDQYCEENG AISTKNLSLK NSTGSISFEG NKSSATGKKG
 201 GAICATGTV DITMNTAPTLF SNNIAEAAGG AINSTGNCTI TGNTSLVFSE
 251 NSVTATAGNG GALSGDADVT ISGNQSVTF SGNQAVANGGA IYAKKLTLAS
 301 GGGGVSPFLT IIVQGTAGN GGAISILAAG ECSSLAEAGD ITFNGNAIVA
 351 TTPQTTRKNS IDIGSTAKIT NLRAISGHSI PFYDPITANT AADSTDITLNL
 401 NKADAGNSTD YSGSIVFSGE KLSEDEAKVA DNLSTLTKQP VTLTAGNLVL
 451 KRGVTLDTKG FTQTAGSSVI MDAGTTLKAS TEEVTLTGLS IPVDSLGEKG
 501 KVVIAASAAS KVALSGPIL LLDNQGNAYE NHDLGKTQDF SFVQLSALGT

5 551 ATTTDVPVAVP TVATPPTHYGY QGTWGMTWVD DTASTPKTKT ATLAWTNTGY
 601 LPNPERQGGL VPNSLWGSFS DIQAIQGVIE RSALTLCSDR GFWAAGVANF
 651 LDKDKKGEKR KYRHKSGGYA IGGAAQTCSE NLISFAFCQL FGSDKDFLVA
 701 KNHDTYAGA FYIQHITECS GFIGCLLDKL PGWSWHPKV LEGQLAYSHV
 751 SNDLKTKYTA YPEVKGSGWN NAFNMMLGAS SHSYPEYLHC FDTYAPYIKL
 801 NLTYIRQDSF SEKGTEGRSF DDSNLFNLSL PIGVKFEKFS DCNDFSVDLT
 851 LSYVPDLIRN DPCKTTALVI SGASWETYAN NLRQALQVR AGSHYAFSPM
 901 FEVLGQFVFE VRGSSRIYNV DLGGKFQF*

A predicted signal peptide is highlighted.

10 The cp0010 nucleotide sequence <SEQ ID 96> is:

1 ATGAAATCGC AATTTTCCTG GTTAGTGCTC TCTTCGACAT TGGCATGTTT
 51 TACTAGTTGT TCCACTGTTT TTGCTGCAAC TGCTGAAAAT ATAGGCCCCCT
 101 CTGATAGCTT TGACGGAAGT ACTAACACAG GCACCTATAC TCCTAAAAAT
 15 151 ACGACTACTG GAATAGACTA TACTCTGACA GGAGATATAA CTCTGCAAAA
 201 CCTTGGGGAT TCGGCAGCTT TAACGAAGGG TTGTTTTTCT GACACTACGG
 251 AATCTTTAAG CTTTGCCGGT AAGGGGTACT CACTTTCTTT TTTAAATATT
 301 AAGTCTAGTG CTGAAGGCGC AGCACTTTCT GTTACAACCTG ATAAAAATCT
 351 TCGCTAACA GGATTTTCGA GTCTTACTTT CTAGCGGCC CCATCATCGG
 401 TAATCACAAC CCCCTCAGGA AAAGGTGCAG TTAATGTGG AGGGGATCTT
 451 ACATTGATA ACAATGGAAC TATTTTATTT AAACAAGATT ACTGTGAGGA
 501 AAATGGCGGA GCCATTCTTA CCAAGAATCT TTCTTTGAAA AACAGCACGG
 551 GATCGATTTC TTTTGAAGGG AATAAATCGA GCGCAACAGG GAAAAAGGT
 601 GGGGCTATTT GTGCTACTGG TACTGTAGAT ATTACAAATA ATACGGCTCC
 25 651 TACCTCTTTC TCGAACAATA TTGCTGAAGC TGCAGGTGGA GCTATAAATA
 701 GCACAGGAAA CTGTACAATT ACAGGGAATA CGTCTCTTGT ATTTTCTGAA
 751 AATAGTGTGA CAGCGACCGC AGGAAATGGA GGAGCTCTTT CTGGAGATGC
 801 CGATGTTACC ATATCTGGGA ATCAGAGTGT AACTTTCTCA GGAACCAAG
 851 CTGTAGCTAA TGGCGGAGCC ATTTATGCTA AGAAGCTTAC ACTGGCTTCC
 901 GGGGGGGGGG GGGTATCTCC TTTTCTAACA ATAAATAGTCC AAGGTACCAC
 30 951 TGCAGGTAAT GGTGGAGCCA TTTCTATACT GGCAGCTGGA GAGTGTAGTC
 1001 TTTTCAAGCA AGCAGGGGAC ATTACCTTCA ATGGGAATGC CATTGTTGCA
 1051 ACTACACCAC AAATACAAA AAGAAATCTT ATTGACATAG GATCTACTGC
 1101 AAAGATCAGC AATTTACGTG CAATATCTGG GCATAGCATC TTTTCTTACG
 35 1151 ATCCGATTAC TGCTAATACG GCTGCGGATT CTACAGATAC TTTAAATCTC
 1201 AATAAGGCTG ATGCAAGTAA TAGTACAGAT TATAGTGGGT CGATTGTTTT
 1251 TTCTGGTGAA AAGCTCTCTG AAGATGAAGC AAAAGTTGCA GACAACCTCA
 1301 CTCTACGCT GAAGCAGCCT GTAACCTTAA CTGCAGGAAA TTTAGTACTT
 1351 AAACGTGGTG TCACTCTCGA TACGAAAGGC TTTACTCAGA CCGCGGGTTC
 40 1401 CTCTGTTATT ATGGATGCGG GCACAACGTT AAAAGCAAGT ACAGAGGAGG
 1451 TCACTTTAAC AGGTCTTTCC ATTCTCTGTAG ACTCTTTAGG CGAGGGTAAG
 1501 AAAGTTGTAA TTGCTGCTTC TGCAGCAAGT AAAAATGTAG CCCTTAGTGG
 1551 TCCGATTCTT CTTTTGGATA ACCAAGGGAA TGCCTATGAA AATCACGACT
 1601 TAGGAAAAAC TCAAGACTTT TCATTTGTGC AGCTCTCTGC TCTGGGTACT
 45 1651 GCAACAACTA CAGATGTTCC AGCGGTTCTT ACAGTAGCAA CTCTACGCA
 1701 CTATGGGTAT CAAGGTACTT GGGGAATGAC TTGGGTTGAT GATACCGCAA
 1751 GCACTCCAAA GACTAAGACA GCGACATTAG CTTGGACCAA TACAGGCTAC
 1801 CTTCCGAATC CTGAGCGTCA AGGACCTTTA GTTCTTAATA GCCTTTGGGG
 1851 ATCTTTTTC A GACATCCAAG CGATTCAAGG TGTCATAGAG AGAAGTGCTT
 50 1901 TGACTCTTTG TTCAGATCGA GGCTTCTGGG CTGCGGGAGT CGCCAATTTC
 1951 TTAGATAAAG ATAAGAAAGG GGAAAAACGC AAATACCGTC ATAAATCTGG
 2001 TGGATATGCT ATCGGAGGTG CAGCGCAAAC TTGTTCTGAA AACTTAATTA
 2051 GCTTTGCTTT TTGCCAATCT TTTGGTAGCG ATAAAGATTT CTAGTCGCT
 2101 AAAAATCATA CTGATACCTA TGCAGGAGCC TTCTATATCC AACACATTAC
 55 2151 AGAATGTAGT GGGTTCATAG GTTGTCTCTT AGATAAACTT CCTGGCTCTT
 2201 GGAGTCATAA ACCCTCGTT TTAGAAGGGC AGCTCGCTTA TAGCCACGTC
 2251 AGTAATGATC TGAAGACAAA GTATACTGCG TATCCTGAGG TGAAAGGTTT
 2301 TTGGGGGAAT AATGCTTTTA ACATGATGTT GGGAGCTTCT TCTCATCTCT
 2351 ATCTTGAATA CCTGCATTGT TTTGATACCT ATGCTCCATA CATCAAATG
 40 2401 AATCTGACCT ATATACGTCA GGACAGCTTC TCGGAGAAAG GTACAGAAGG
 2451 AAGATCTTTT GATGACAGCA ACCTCTTCAA TTTATCTTTG CCTATAGGGG
 2501 TGAAGTTTGA GAAGTCTCTT GATTGTAATG ACTTTTCTTA TGATCTGACT
 2551 TTATCCTATG TTCCTGATCT TATCCGCAAT GATCCCAAAT GCACACAGC
 2601 ACTTGTAATC AGCGGAGCCT CTTGGGAAAC TTATGCCAAT AACTTAGCAC
 2651 GACAGGCCCT GCAAGTGCCT GCAGGCAGTC ACTACGCTT CTCTCCTATG
 65 2701 TTTGAAGTGC TCGGCCAGTT TGTCTTTGAA GTTCGTGGAT CCTCAGGAT

2751 TTATAATGTA GATCTTGGGG GTAAGTTCCA ATTCTAG

The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 48A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 48B) and for FACS analysis (Figure 48C). A his-tagged protein was also expressed.

The cp0010 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0010 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

10 Example 49

The following *C.pneumoniae* protein (PID 4376296) was expressed <SEQ ID 97; cp6296>:

```

1  MEEVSEYLQQ VENQLESCSK RLTKMETFAL GVRLEAKEEI ESILSDVWN
51  RFEVLCRDIE DMLSRVEEIE RMLRMAELPL LPIKEALTKA FVQHNSCKEK
101 LTKVEPYFKE SPAYLTSEER LQSLNQTLQR AYKESQKVSG LESEVRACRE
151 QLKDQVRQFE TQGVSLIKEE ILFVTSTFRT KFSYHSFRLH VPCMRLYEEY
201 YDDIDLERTR ARWMAMSERV RDAFQAFQEM LKEGLVEEAQ ALRETEYWLY
251 REERKSKKKH*
```

The cp6296 nucleotide sequence <SEQ ID 98> is:

```

1  ATGGAGGAGG TGTCTGAGTA TCTTCAGCAA GTAGAAAATC AGTTGGAATC
51  CTGTTCCAAG CGATTAACCA AGATGGAAAC TTTTGCCTTA GGTGTGAGGT
101 TGGAAGCTAA AGAAGAGATA GAGTCTATCA TACTTTCTGA TGTAAGTAAC
151 CGTTTTGAGG TTTTATGTAG AGATATTGAA GATATGCTAT CTCGAGTCGA
201 GGAGATAGAG CGGATGTTAC GTATGGCGGA GCTTCCTCTA CTTCTCTATAA
251 AAGAAGCGCT TACCAAGGCT TTTGTACAAC ATAACAGCTG TAAAGAGAAG
301 TTAACCAAGG TAGAGCCTTA CTTTAAAGAG AGCCCTGCAT ATCTAAGTAG
351 TGAAGAGCGA TTGCAGAGTT TGAATCAGAC TTTACAACGT GCGTACAAAG
401 AGTCCCAAAA GGTTCAGGT TTAGAATCGG AAGTGAGAGC CTGTCGAGAG
451 CAGCTTAAAG ATCAAGTAAG ACAGTTTGAA ACTCAAGGAG TGAGCTTGAT
501 AAAAGAAGAG ATTCTCTTTG TGACTAGTAC CTTTAGAAGT AAATTTAGCT
30  551 ATCATTCATT TCGATTACAT GTTCTTGCA TGAGGTTGTA TGAGGAGTAT
601 TATGATGACA TTGATCTAGA GAGAACTCGA GCTCGATGGA TGGCGATGTC
651 TGAGAGGTAT AGAGATGCTT TTCAGGCATT CCAGGAGATG TTGAAGGAAG
701 GCCTAGTTGA AGAAGCTCAG GCTCTTAGAG AAACCGAGTA CTGGTTATAT
751 CGAGAGGAGA GAAAGAGTAA AAAGAAACAT TGA
```

35 The PSORT algorithm predicts a cytoplasmic location (0.523).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 49A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 49B) and for FACS analysis (Figure 49C). A his-tagged protein was also expressed.

These experiments show that cp6296 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 50

The following *C.pneumoniae* protein (PID 4376664) was expressed <SEQ ID 99; cp6664>:

```

1  MVLFAQASG RNRVKADAIV LPFWHFKDAK NAASFEEFE PSYLPALENF
51  QGKTGEIELL YSSPKAKEKR IVLLGLGKNE ELTSDVVFQT YATLTVLRK
101 AKCSTVNIL PTISELRISA EEFLVGLSSG ILSLNYDYPR YNKVDRNLET
```

5 151 PLSKVTVIGI VPKMADAIFR KEAAIFEGVY LTRDLVNRNA DEITPKKLAE
 201 VALNLGKEFP SIDTRVLGKD AIAKEKMGLL LAVSKGSCVD PHFIVVRYQG
 251 RPKSKDHTVL IKGVTTFDSG GLDLKPGKSM LTMKEDMAGG ATVLGILSAL
 301 AVLELPINVT GIIPATENAI DGASYKMGDV YVGMSSGLSVE ICSTDAEGRL
 351 ILADAITYAL KYCKPTRIID FATLTGAMVV SLGEEVAGFF SNNDVLAEDL
 401 LEASAETSEP LWRLPLVKKY DKTLLHSDIAD MKNLGSNRAG AITAALFLQR
 451 FLEESSVAWA HLDIAGTAYH EKEEDRYPKY ASGFGVRSIL YYLENSLSK*

The cp6664 nucleotide sequence <SEQ ID 100> is:

10 1 GTGGTTTAT TTCATGCTCA AGCCTCTGGG CGTAATCGTG TTAAGGCAGA
 51 TGCTATAGTC CTGCCCTTTT GGCATTTTAA GGATGCAAAA AATGCAGCTT
 101 CTTTGAAGC CGAGTTTGAA CCCTCGTATC TCCCCGCTTT AGAAAACTTT
 151 CAAGGAAAAA CCGGGGAGAT TGAACCTCTT TATAGTAGTC CTAAAGCTAA
 201 GGAAAAACGC ATTGTCTCTT TAGGCTTAGG GAAAAATGAA GAGCTCACCT
 251 CTGATGTTGT TTTCCAAACC TATGCGACAC TAACTCGTGT CTTACGTAAA
 301 GCAAAGTGTT CCACAGTCAA TATCATCTTA CCTACAATTT CTGAATTGCG
 351 GCTTTCTGCC GAAGAATTCT TAGTGGGGTT GTCCCTCAGGA ATTTTGTCTAT
 401 TAAACTATGA CTACCCACGT TATAATAAGG TAGATCGTAA TCTTGAAACT
 451 CCTCTTTCTA AAGTCACGGT TATCGGTATC GTTCCCAAAA TGGCGGATGC
 501 TATCTTTAGG AAAGAAGCAG CCATTTCGA AGGCGTATAT CTCACTCGAG
 551 ATCTTGTAAG CAGGAATGCT GATGAAATTA CCCCTAAGAA ATTGGCAGAG
 601 GTTGCTCTGA ATCTGGGAAA AGACTTCCCT AGTATTGATA CTAAGGTCTT
 651 GGGAAAAGAT GCCATCGCCA AAGAGAAAAT GGGACTCCTA TTGGCTGTTT
 701 CCAAGGGTTC TTGTGTGGAT CCACACTTTA TCGTTGTCCG TTATCAAGGA
 25 751 CGTCTAAGT CTAAAGATCA CACCGTCTTG ATAGGGAAAG GGGTCACTTT
 801 TGACTCTGGA GGTTTAGACC TCAAGCCTGG AAAATCCATG CTTACTATGA
 851 AAGAAGACAT GGCAGGTGGG GCTACAGTCC TCGGGATTCT CTCGGCGTTA
 901 GCAGTTTTCG AGCTTCCTAT AAATGTCACG GGGATCATTC CTGCTACAGA
 951 GAATGCTATC GATGGCGCCT CCTATAAAAT GGGAGATGTC TATGTAGGAA
 30 1001 TGTCGGGGCT TTCTGTTGAG ATTTGTAGTA CCGATGCTGA GGGACGTCTT
 1051 ATCCTCGCTG ATGCCATTAC ATATGCTTTA AAATATTGTA AACCGACACG
 1101 TATTATAGAT TTTGCAACTC TAACAGGAGC TATGTTAGTC TCTCTAGGAG
 1151 AAGAGGTGTC AGGTTTCTTT TCCAATAACG ATGTTTTCAG TGAAGATCTT
 1201 TTAGAGGCGT CAGCCGAAAC CTCCGAGCCG TTATGGAGAC TTCTCTAGT
 1251 TAAGAAGTAT GATAAAACAT TGCAATTCTGA TATTGCTGAT ATGAAAATC
 35 1301 TAGGCAGTAA CCGTGCAGG GCTATTACAG CAGCATTTAT CTTGCAGAGA
 1351 TTTTGGGAAG AATCTICGGT AGCTTGGGCA CATCTTGATA TTGCAGGTAC
 1401 TGCATATCAT GAAAAAGAAG AAGACCGTTA TCCAAAATAT GCTTCAGGTT
 1451 TTGGTGTTCG TTCTATTCTT TATTACTTAG AAAATAGTCT TTCTAAGTAG

The PSORT algorithm predicts an inner membrane location (0.268).

- 40 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 50A), as a his-tagged protein, and as a GST/His fusion. The proteins were used to immunise mice, whose sera were used in Western blot Western blot (50B) and FACS (50C) analyses.

The cp6664 protein was also identified in the 2D-PAGE experiment (Cpn0385) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 45 These experiments show that cp6664 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 51

The following *C.pneumoniae* protein (PID 4376696) was expressed <SEQ ID 101; cp6696>:

50 1 MTLIFVIIIV WCNALIKLC VIMGLQSRLO HCIEVSQNSN FDSQVKQFIY
 51 ACQDKTLRQS VLKIFRYHPL LKIHDIARAV YLLMALEBGE DLGLSFLNVQ
 101 QYPSGAVELF SCGGFPWKGL PYPAEHAEFG LLLLQIAEFY EESQAYVSKM
 151 SHFQQALFDH QGSVFPSSLWS QENSRLLEBK TTLSQSFLFQ LGMQIHPEYS
 201 LEDPALGFWM QTRSSSAFV AASGCQSSLG AYSSGDVGVI AYGPCSGDIS
 251 DCYYFGCCGI AKEFVCQKSH QTTEISPLTS TGKPHPRNTG FSYLRDSYVH
 55 301 LPIRCKITIS DKQYRVHAAL AEATSAMTFS IFCKGKNCQV VDGPRLRSCS

351 LDSYKGPND IMILGENDAI NIVSASPYME IFALQGKEKF WNADFLINIP
401 YKEGVMLIF EKKVTSEKGR FFTKMN*

A predicted signal peptide is highlighted.

The cp6696 nucleotide sequence <SEQ ID 102> is:

```

5      1 TTGACTCTAA TTTTGTAT TATTATCGTT TGGTGCAATG CTTTCTGAT
      51 CAAATGTGTC GTGATAATGG GGCTGCAATC CAGGTTACAA CATTGTATAG
101    AAGTGTCCCA GAATTCGAAC TTTGATTAC AAGTAAACA GTTTATCTAT
151    GCGTGCCAAG ATAAGACATT AAGGCAGTCT GTACTCAAGA TTTCCGCTA
201    CCATCCTTTA CTAAAAATTC ATGATATTGC TCGGGCCGTC TATCTTTTGA
10      251 TGGCCTTAGA AGAAGGCGAG GATTTAGGCT TAAGCTTTT AAATGTACAG
      301 CAGTACCCCT CAGGTGCTGT AGAACTGTTT TCTGTGGGG GATTTCCTTG
      351 GAAAGGATTA CCTTATCCTG CAGAACATGC GGAATTGGG CTACTCCTGT
      401 TACAGATCGC AGAGTTTAT GAAGAGAGTC AGGCATACGT CTCTAAATG
15      451 AGTCATTTTC AACAGGCACT CTTTGATCAC CAAGGAGCG TCTTCCCTC
      501 TCTCTGGAGC CAGGAGAACT CTCGACTCCT AAAAGAAAAG ACAACTCTTA
      551 GCCAATCGTT TCTCTCCAA TTAGGAATGC AAATCACCC AGAATACAGT
      601 CTTGAGGATC CTGCACTAGG GTTCTGGATG CAAAGAACGC GTTCTTCATC
15      651 CGCTTTTGTA GCCGCTTCAG GATGTCAAAG TAGCTTGGGA GCGTATTCCT
      701 CAGGGGATGT CGGTGTTATC GCTTATGGAC CTTGCTCTGG AGACATTAGT
20      751 GATTGTTATT ATTTTGGATG TTGTGGAATC GCTAAAGAGT TCGTGTGCCA
      801 AAAATCTCAC CAACTACAG AGATTTCTTT TCTCACCTCT ACAGGAAAGC
      851 CTCATCCCAAG AAATACGGGA TTTTCTTACC TTCGAGATTC CTATGTACAT
      901 CTGCCGATCC GCTGTAAGAT CACTATTTC GACAAGCAAT ATCGCGTGCA
25      951 CGCTGCGTTG GCTGAGGCCA CCTCTGCCAT GACGTTTTCT ATTTCTGTA
      1001 AGGGGAAGAA TTGTCAGGTT GTTGACGGCC CTCGCTTGCG CTCTGTTC
      1051 CTAGATTCTT ATAAAGGTCC CGGAAACGAC ATTATGATTC TTGGGGAAAA
      1101 TGACGCAATC AACATTGTTT CTGCAAGTCC CTATATGGAA ATTTTGTCTT
      1151 TGCAAGGCAA AGAAAAATTT TGAATGCAG ACTTTTGAT TAATATCTCT
30      1201 TACAAAGAAG AGGCGTCAT GTTAATTTT GAAAAAAG TGACCTCTGA
      1251 GAAAGGAAGA TTCTTTACGA AGATGAATTA A

```

The PSORT algorithm predicts an inner membrane location (0.463).

The protein was expressed in *E. coli* and purified as a GST-fusion product, as shown in Figure 51A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 51B) and for FACS analysis (Figure 51C). A his-tagged protein was also expressed.

35 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6696 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 52

40 The following *C. pneumoniae* protein (PID 4376790) was expressed <SEQ ID 103; cp6790>:

```

1      MSEHKKSSKI IGIDLGTNS CVSVMEGGQA KVITSSEGTR TTPSIVAFKG
51     NEKLVGIPAK RQAVTNPEKT LGSTKRFIGR KYSEVASEIQ TVPYTWTSGS
101    KGDVFEVDG KQYTPBEGA QILMKMKETA BAYLGETVTE AVITVPAYFN
15      151 DSQRASKDA GRIAGLDVGR IPEPTAAAL AYGIDKVGDK KIAVFDLGGG
      201 TFDISILEIG DGVFEVLSTN GDTLLGGDDF DEVIKWMIE EFKKQEGIDL
      251 SKDNMALQRL KDAAEKAKIE LSGVSSTEIN QPFITMDAQQ PKHLALTLTR
      301 AQFEKLAASL IERTKSPCIK ALSDAKLSAK DIDDVLLVGG MSRMPAVQET
      351 VKELFGKEPN KGVNPDEVVA IGAAIQGGVL GGEVKDVLLE DVIPLSLGIE
15      401 TLGGVMTTLV ERNTTIPTQK KQIFSTAADN QPAVTIVVLQ GERPMADNK
      451 EIGRFDLTDI PPAPRGHPQI EVSFDIDANG IFHVSADVA SGKEQKIRIE
      501 ASSGLQEDEI QRMVRDAEIN KEEDKKRREA SDAKNEADSM IFRAEKA IKD
      551 YKEQIPETLV KEIBERIENV RNALKDDAPI EKIKEVTEDI SKHMOKIGES
      601 MQSQSASAAA SSAANAKGGP NINTEDLKKH SFSTKPPSNN GSSEDHIEEA

```


651 DVEIIDNDDK*

The cp6790 nucleotide sequence <SEQ ID 104> is:

```

      1 ATGAGTGAAC ACAAAAAATC AAGCAAAATT ATAGGTATAG ACTTAGGCAC
      51 AACAAACTCC TGCCTATCTG TTATGGAAGG AGGACAAGCT AAAGTAATTA
      101 CATCATCCGA AGGAACAAGA ACCACGCCAT CGATCGTTGC CTTCAAAGGT
      151 AATGAGAAAT TAGTGGGGAT TCCAGCAAAA CGTCAAGCAG TGACAAATCC
      201 AGAAAAAATC CTCGGCTCTA CAAAACGCTT TATTGGCCGT AAGTACTCTG
      251 AAGTAGCTTC GGAAATCCAA ACCGTTCTCT ATACAGTCAC CTCCGGATCT
      301 AAAGGTGATG CCGTTTTCGA AGTTGATGGC AAACAATACA CTCCAGAAGA
      351 AATTGGCGCA CAAATCTTAA TGAAAATGAA AGAGACAGCA GAAGCTTATC
      401 TAGGCGAAAC TGTCACAGAA GCAGTGATCA CCGTCCCCGC ATACTTCAAT
      451 GATTCTCAAC GAGCATCCAC AAAAGATGCT GGACGCATTG CAGGTCTAGA
      501 TGTA AACGT ATCATTCCAG AACCTACCGC AGCAGCTCTT GCCTACGGAA
      551 TCGATAAAGT CCGTGATAAA AAAATCGCTG TCTTCGACCT TGGTGGAGGA
      601 ACTTTTGATA TCTCCATCCT AGAAATCGGT GATGGCGTCT TCGAAGTTCT
      651 ATCTACAAAT GGAGATACTC TCCTCGGTGG AGACGACTTT GATGAAGTCA
      701 TTATCAAAATG GATGATCGAA GAATCAAAA AACAAGAAGG CATTGATCTT
      751 AGCAAAGATA ATATGGCCTT ACAAAGACTT AAAGATGCTG CTGAGAAAGC
      801 AAAAAATAGAA CTTTCAAGGAG TCTCTCCAC AGAAATCAAT CAGCCATTCA
      851 TCACAATGGA TGCACAAGGA CCTAAACACC TTGCATTGAC ACTCACACGT
      901 GCGCAATTCG AGAAACTCGC AGCCTCTCTA ATCGAAAGAA CAAATCTCC
      951 ATGCATCAAA GCACTCAGTG ACGCAAAACT TTCCGCTAAG GATATCGATG
      1001 ATGTTCTCTT AGTTGGAGGT ATGTCAAGAA TGCCCCGAGT GCAAGAAACT
      1051 GTAAAAGAAC TCTTCGGCAA AGAGCCTAAT AAAGGAGTCA ACCCCGACGA
      1101 AGTTGTTGCT ATTGGAGCCG CAATTCAAGG TGGTGTCTT GCGGAGAGAAG
      1151 TTAAGGATGT TCTACTTCTA GACGTATACC CCCTATCTCT GGGTATCGAA
      1201 ACTCTAGGAG GCGTCATGAC GACTCTGGTA GAGAGAAATA CTACAATCCC
      1251 TACACAGAAA AAACAAATCT TCTCCACAGC TGCTGATAAC CAGCCTGCGG
      1301 TTACCATCGT AGTTCTCCAA GGAGAGCGTC CCATGGCCAA AGATAACAAG
      1351 GAAATCGGAA GATTGATCTT TACAGATATC CCTCCGGCTC CTCGAGGCCA
      1401 TCCTCAAATC GAAGTCTCCT TCGATATCGA TGCAAACGGA ATTTTCCATG
      1451 TCTCAGCTAA AGATGTTGCC AGCGGTAAAG AACAGAAAT TCGTATCGAA
      1501 GCAAGCTCAG GACTTCAAGA AGATGAAATC CAAAGAATGG TTCGAGATGC
      1551 CGAAATTAAT AAGGAAGAAG ATAAAAACG TCGTGAAGCT TCAGATGCTA
      1601 AAAATGAAGC CGATAGCATG ATCTTCAGAG CCGAAAAAGC TATTAAAGAT
      1651 TATAAGGAGC AAATTCCTGA AACTTTAGTT AAAGAAATCG AAGAGCGAAT
      1701 CGAAAACGTG CGCAACGCAC TCAAAGATGA CGCTCCTATT GAAAAAATTA
      1751 AAGAGGTTAC TGAAGACCTA AGCAAGCATA TGCAAAAAAT TGGAGAGTCT
      1801 ATGCAATCGC AGTCTGCATC AGCAGCAGCA TCATCGGCAG CCAATGCTAA
      1851 AGGTGACCTT AACATCAATA CAGAAGATTT GAAAAACAT AGTTTCAGTA
      1901 CGAAGCCTCC TTCAAATAAC GGTCTTCAG AAGACCATA CAGAAGAAGCT
      1951 GATGTAGAAA TTATTGATAA CGACGATAAG TAA

```

The PSORT algorithm predicts an inner membrane location (0.151).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 52A) and a his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 52B) and FACS (Figure 52C) analyses.

The cp6790 protein was also identified in the 2D-PAGE experiment (Cpn0503).

These experiments show that cp6790 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 53The following *C.pneumoniae* protein (PID 4376878) was expressed <SEQ ID 105; cp6878>:

```

      1 MNVPDSKNLH PPAYELLEIK ARITQSYKEA SAILTAIPDG ILLLSETGHF
      51 LICNSQAREI LGIDENLEIL NRSFTDVLDP TCLGFSIQEA LESLKVPKTL
      101 RLSLCKESKE KEVELFIRKN EISGYLFIQI RDRSDYKQLE NAIERYKNIA
      151 ELGKMTATLA HEIRNPLSGI VGFASILKKE ISSPRHQRL SSIISGTRSL
      201 NNLVSSMLEY TKSQPLNLKI INLQDFPSSL IPLLVSFPN CKFVREGAQP

```


251 LFRSIDPDRM NSVWNLVKN AVETGNSPIT LFLHTSGDIS VTNPGTIPSE
 301 IMDKLFPTFF TTKREGNGLG LAEAQKIIRL HGGDIQLKTS DSAVSFFIII
 351 PELLAALPKE RAAS*

The cp6878 nucleotide sequence <SEQ ID 106> is:

5 1 ATGAACGTCC CTGATTCCAA GAACCTCCAT CCTCCTGCAT ACGAACTCCT
 51 AGAGATCAAG GCTCGCATCA CACAATCTTA TAAAGAAGCG AGTGCTATAC
 101 TGACAGCGAT TCCTGATGGT ATCCTATTAC TTCTGAAAC AGGACACTTT
 151 CTTATCTGCA ATTCACAAGC ACGTGAATTT CTAGGAATTG ATGAAAATCT
 201 AGAAATTTCTT AATAGATCCT TTACCGATGT TCTCCCGAT ACGTGTCTTG
 10 251 GATTTTCTAT TCAAGAGGCT CTTGAATCTC TAAAGTCCC TAAACTCTCT
 301 AGACTCTCTC TCTGTAAAGA ATCTAAAGAA AAAGAAGTGG AACTCTTCAT
 351 CCGTAAAAAC GAGATCAGTG GATACCTGTT TATCCAAATC CGCGATCGGT
 401 CCGACTATAA ACAACTAGAA AACGCTATAG AAAGATATAA AAATATCGCA
 451 GAACCTGGGA AATGACGGC TACCCTAGCT CACGAAATCC GCAATCCGCT
 15 501 AAGTGGAAATC GTTGGATTG CCTCTATCCT AAAGAAAGAG ATTTCCTCTC
 551 CTCGCCACCA ACGAATGCTC TCCTCAATCA TCTCCGGCAC AAGGTCTCTA
 601 AATAACCTTG TCTCTCTAT GTTAGAATAT ACAAATCAC AACCGTTGAA
 651 CATAAAGATT ATAAATTTAC AAGACTTCTT CTCTTCTCTT ATCCCTCTGC
 701 TCTCCGTCTC TTTCCCGAAT TGCAAGTTTG TAAGAGAGGG CGCACAACCT
 20 751 CTATTCAGAT CTATAGATCC TGATCGGATG AACAGTGTG TTTGGAACCT
 801 AGTGAAAAAT GCTGTAGAAA CAGGGAACCT TCCGATCACT CTGACCCTGC
 851 ATACATCGGG AGACATCTCG GTAACGAACC CCGGAACGAT TCCTTCCGAG
 901 ATCATGGACA AGCTCTTCAC TCCTATCTTC ACAACAAAGA GAGAGGGAAA
 951 TGGTTTGGGA CTTGCTGAAG CTCAAAAAT TATAAGACTC CATGGAGGAG
 25 1001 ATATCCAATT AAAACAAGC GACTCCGCG TTAGCTTCT CATAATCATC
 1051 CCCGAAC TTC TAGCGGCCCT ACCCAAAGAA AGAGCCGCTA G

The PSORT algorithm predicts an inner membrane location (0.204).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 53A) and as a GST-fusion product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 53B) and for FACS analysis.

These experiments show that cp6878 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 54

The following *C.pneumoniae* protein (PID 4377224) was expressed <SEQ ID 107; cp7224>:

35 1 MMKIKRIVAL AVGGSGGHIV PALSVKEAFS REGIDVLLLG KGLKNHPSLQ
 51 QGISYREIPS GLPTVLNPIK IMSRTLSLCS GYLKARKELK IFDPDLVIGF
 101 GSYHSLPVLL AGLSHKIPLF LHEQNLVPGK VNQLFSRYAR GIGVNFSPVT
 151 KHFRCPAEV FLPKRSFSLG SPMKRCCTNH TPTICVVGGS QGAQILNLCV
 201 PQALVKLVNK YPNMYVHHIV GPKSDVMKVQ HVYNRGEVLC CVKPFEEQLL
 40 251 DVLLAADLVI SRAGATILEE ILWAKVPGIL IPYPGAYGHQ EVNAKFFVDV
 301 LEGGTMILEK ELTEKLLVEK VTFALDSHNR EKQRNSLAAY SQQRSTKTFH
 351 AFICECL*

The cp7224 nucleotide sequence <SEQ ID 108> is:

45 1 ATGATGAAGA AAATTCGAAA AGTAGCCTTG GCTGTAGGAG GTTCAGGAGG
 51 CCACATTGTC CCAGCTCTCT CGGTAAAGGA AGCTTTTCTT CGTGAAGGAA
 101 TAGACGTATT ACTACTAGGG AAAGGTCTCA AGAACCATCC TTCTTTGCAA
 151 CAGGGAATCA GCTATCGGGA AATCCCCCTCA GGACTTCCTA CAGTCCCTAA
 201 TCCCATAAAG ATCATGAGCA GGACCCCTTC TCTATGTTCA GGATACCTGA
 251 AAGCAAGAAA GGAACTTAAA ATTTTGTACC CTGACCTGGT CATAGGATTT
 50 301 GGGAGCTACC ACTCTCTTCC CGTGTGCTC GCAGGACTGT CCCATAAAAT
 351 TCCCTTATTT CTACACGAAC AAAATCTAGT TCCTGGAAAA GTAAATCAAT
 401 TGTTTTCCTG CTATGCTCGA GGTATTGGAG TGAATTTCTC CCCCGTTACT
 451 AAACACTTCC GCTGCCCGGC AGAAGAGGTC TTCCTTCCTA AACGAAGCTT
 501 CTCCTTAGGA AGCCCTATGA TGAAGCGATG TACAAATCAT ACCCTACAA
 55 551 TCTGTGTTGT TGGAGGTTCT CAGGGAGCAC AGATATTAAA TACTTGTGTT
 601 CCCCAAGCTC TTGTCAAGCT AGTCAATAAG TACCCAAATA TGTACGTCCA

5
 651 TCATATTGTA GGACCTAAAA GTGATGTTAT GAAGGTGCAA CATGTTTACA
 701 ATCGTGGAGA GGTCTCTGTC TGTGTGAAGC CGTTCGAAGA GCAACTCCTA
 751 GATGTCTTGC TTGCCGCAGA TTTGGTCATC AGTAGGGCAG GAGCCACAAT
 801 TTTAGAAGAA ATTCTTTGGG CAAAAGTTCC CGGAATTTTA ATTCCTTATC
 851 CAGGAGCTTA TGGACATCAG GAAGTTAATG CTAAATTCCT TGTAGACGTC
 901 TTAGAAGGGG GAACTATGAT CCTAGAAAAA GAATTAACAG AGAAGCTATT
 951 AGTAGAAAAA GTAACGTTTG CTTTAGACTC CCATAACAGA GAAAAACAAC
 1001 GCAATTCCCT AGCGGCGTAT AGTCAGCAAA GGTCAACAAA AACATTCCAT
 1051 GCATTCATTT GTGAATGCTT ATAG

10 The PSORT algorithm predicts an inner membrane location (0.164).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 54A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 54B) and for FACS analysis (Figure 54C). A his-tagged protein was also expressed.

15 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7224 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 55

The following *C.pneumoniae* protein (PID 4377140) was expressed <SEQ ID 109; cp7140>:

20
 1 MVRRSISFCL FFLMTLLCCT SCNSRSLIVH GLPGREANEI VVLLVSKGVA
 51 AQKLPQAAAA TAGAATEQMW DIAVPSAQIT EALAILNQAG LPRMKGTSLL
 101 DLFAKQGLVP SELQEKIRYQ EGLSEQMAST IRKMDGVVDA SVQISFTTEN
 151 EDNLPLTASV YIKHRGVLDN PNSIMVSKIK RLIASAVPGL VPENVSUVSD
 201 RAAYSDITIN GPWGLTEEID YVSVWGIIIA KSSLTKFRLLI FVVLILILFV
 25
 251 ISCGLLWVIW KTHTLIMTMG GTKGFFNPTP YTKNALEAKK AEGAAADKEK
 301 KEDADSQGES KNAETSDKDS SDKDAPEGSN BIEGA*

A predicted signal peptide is highlighted.

The cp7140 nucleotide sequence <SEQ ID 110> is:

30
 1 ATGGTTCGTC GATCTATTTT TTTTGTCTTG TTCTTTCTAA TGACATTGCT
 51 GTGCTGTACA AGCTGTAACA GCAGGTCTCT AATTGTGCAC GGTCTTCCTG
 101 GCAGAGAAGC GAATGAGATT GTGGTGCTTT TGGTAAGCAA AGGGGTGGCT
 151 GCACAAAAAT TGCCTCAAGC TGCAGCGGCT ACAGCCGGAG CAGCTACTGA
 201 GCAAAATGTGG GATATCGCGG TTCCGTCAGC ACAATACACA GAGGCCCTTG
 251 CCATTCTAAA TCAAGCGGGT CTTCCACGTA TGAAAGGGAC AAGCCTGTTA
 35
 301 GATCTTTTGT CAAAACAAGG TCTTGTTCCT TCCGAGCTTC AGGAAAAAAT
 351 CCGTTATCAA GAAGGCTTAT CAGAACAGAT GGCCTCTACG ATTAGAAAAA
 401 TGGATGGCGT TGTGATGCC TCAGTACAGA TTTCCTTCAC TACAGAAAAAT
 451 GAAGATAATC TTCTTTTAAC AGCCTCTGTG TATATTAAGC ATCGAGGGGT
 501 TTTGGACAAAT CCGAACAGCA TTATGGTTTC CAAAATTAAG CGCCTTATTG
 40
 551 CAAGTGCTGT TCCAGGACTT GTGCCAGAGA ACGTCTCTGT AGTGAGCGAT
 601 CGCGCAGCTT ATAGTGATAT TACAATTAAT GGTCTTGGG GATTAAACGA
 651 AGAAATCGAT TATGTTTCTG TTTGGGGTAT TATTCTTGCG AAGTCTTCGC
 701 TCACCAAATP CCGTCTCATT TTTTATGTCT TGATTCTCAT TTTATTTGTT
 751 ATTCTTTGTG GTCTCCTTTG GGTCAATTTG AAAACTCATA CTCTCATTAT
 45
 801 GACTATGGGA GGTACAAAAG GGTCTTCAA CCCTACACCA TATACAAAGA
 851 ATGCCTTGGA AGCCAAGAAA GCCGAGGAG CAGCTGCTGA CAAAGAGAAA
 901 AAAGAAGATG CAGATTCACA GGGGGAAGC AAAAATGCGG AAACCAAGTA
 951 TAAAGACTCT AGTGATAAAG ATGCTCCAGA AGGAAGCAAT GAAATTGAGG
 1001 GTGCTTAG

50 The PSORT algorithm predicts an inner membrane location (0.650).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 55A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 55B) and for FACS analysis (Figure 55C). A his-tagged protein was also expressed.

These experiments show that cp7140 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 56

The following *C.pneumoniae* protein (PID 4377306) was expressed <SEQ ID 111; cp7306>:

```

1  MTKQLRSWL AVLVGSSLLA LPLSGQAVGK KESRVSELPQ DVLLEKISGG
51 FSKVATKATP AVVYIESFPK SQAVTHPSPG RRGPYENPFD YFNDEFFNRF
101 FGLPSQREKP QSKEAVRGTG FLVSPDGYIV TNNHVVEDTG KIHVTLHDQ
151 KYPATVIGLD PKTDLAVIKI KSQNLPLYLSF GNSDHLKVGD WATAIGNPFG
201 LQATVTVGVI SAKGRNQLHI ADFEDFIQTD AAINPGNSGG PLLNIDGQVI
251 GVNTAIVSGS GGYIGIGFAI PSLMANRIID QLIRDGQVTR GFLGVTLQPI
301 DAELAACYKL EKVYGALVTD VIKGSPADKA GLKQEDVIA YNGKEVDSLS
15 MFRNAVSLMN PDTRIVLKVV REGKVIEIPV TVSQAPKEDG MSALQVRGIR
401 VQNLTPETAK KLGIAPETKG ILIISVEPGS VAASSGIAPG QLILAVNRQK
451 VSSIEDLNRT LKDSNNENIL LMVSQGDVIR FIALKPEE*
```

A predicted signal peptide is highlighted.

The cp7306 nucleotide sequence <SEQ ID 112> is:

```

20 1  ATGATAACTA AGCAATTGCG TTCGTGGCTA GCTGTACTTG TTGGTTCAAG
51  TCTGCTAGCT CTTCCCTTAT CAGGGCAAGC TGTCGGGAAA AAAGAATCTC
101 GAGTTTCCGA GCTGCCTCAA GACGTTCTTC TTAAAGAGAT CTCGGGAGGG
151 TTTTCTAAGG TCGCTACCAA GGCGACTCCC GCTGTTGTGT ACATAGAAAG
201 TTCCCCAAG AGCCAGGCTG TAACACATCC TTCTCCTGGA CGCCGTGGGC
25 251 CTTATGAAAA TCCTTTTGAT TATTTTAATG ATGAGTTTTC CAATCGTTTT
301 TTTGGTCTAC CTTACAGAG GGAAAAACCT CAAAGTAAAG AGGCGGTTTCG
351 AGGAACAGGT TTCCTAGTAT CTCCAGATGG CTATATTGTG ACTAATAACC
401 ATGTTGTGCA AGATACAGGT AAGATTCACG TAACTCTTCA TGATGGGCAA
451 AAGTACCCAG CAACTGTAAT CGGACTCGAT CCTAAAACAG ACCTTGCACT
30 501 CATTAATAAT AAATCCCAAA ACCTCCCGTA TCTTCTTTT GGAAACTCCG
551 ACCACTTAAA AGTCGGAGAT TGGGCAATTG CAATTGGAAA TCCCTTCGGT
601 CTTCAAGCTA CGGTCACCGT AGGTGTCATC AGTGCTAAAG GAAGAAATCA
651 ACTCCACATT GCAGATTTTG AAGATTTTAT TCAGACAGAT GCTGCGATT
701 ATCCAGGCAA CTCTGGAGGC CCTCTTCTAA ATATTGATGG ACAGGTCATC
35 751 GGTGTTAATA CTGCCATTGT CAGTGGTAGT GGTGGCTATA TTGGAATCGG
801 GTTTGCGATT CCTAGCCTTA TGGCAAAATG AATCATAGAT CAGCTGATTC
851 GTGATGGTCA AGTTACCCGA GGATTCCTAG GAGTGACTTT ACAACCTATA
901 GATGCGGAAC TCGCTGCTTG CTACAAACTC GAAAAGGTTT ATGGCGCTTT
951 AGTCACAGAT GTTGTTAAAG GATCTCCAGC AGATAAAGCA GGGCTAAAC
40 1001 AAGAAGATGT GATCATTGCT TATAATGGGA AAGAAGTCTG TTCACAGT
1051 ATGTTCCGTA ATGCTGTTTC TTTAATGAAT CCAGATACAC GTATTGTTCT
1101 AAAGGTAGTT CGTGAAGGAA AGGTTATCGA AATACCCGTG ACAGTTTCTC
1151 AAGCTCCAAA AGAAGATGGA ATGTCGGCTT TACAGCGTGT GGGAAATCCGT
1201 GTGCAAAACC TAACTCCTGA AACTGCTAAG AAGCTGGGAA TTGCTCCAGA
45 1251 GACTAAAGGC ATTTTGATTA TAAGTGTGTA ACCAGGGTCT GTAGCAGCTT
1301 CTTCAGGAAT TGCTCCCTGGT CAGCTGATCC TTGCTGTGAA TAGACAAAAA
1351 GTATCTTCGA TTGAAGATCT GAATAGAACG TTAAGAGATT CTAACAATGA
1401 GAATATTCTT CTTATGGTTT CTCAAGGAGA TGTTATTGCG TTCATTGCCC
1451 TGAACCTGA AGAATAA
```

50 The PSORT algorithm predicts a periplasmic location (0.923).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 56A) and as a GST-fusion product (Figure 56B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 56C) and for FACS (Figure 56D) analyses.

The cp7306 protein was also identified in the 2D-PAGE experiment (Cpn0979) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7306 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 Example 57

The following *C.pneumoniae* protein (PID 4377132) was expressed <SEQ ID 113; cp7132>:

```

1  MCNSIAMKKQ KRGFVLMELL MSFTLLIAILL GTLGFWYRKI YTVQKQKERI
51  YNFYIEESRA YKQLRTLFSM SLSSSYEEPG SLFSLIFDRG VYRDPKLAGA
101 VRASLHHDTK DQRLELRICN IKDQSYFETQ RLLSHVTHV LSFQRNPDPE
151 KLPETIALTI TREPKAYPPR TLTYQFAVGK*

```

A predicted signal peptide is highlighted.

The cp7132 nucleotide sequence <SEQ ID 114> is:

```

1  ATGTGTAAC TATAGCTAT GAAAAGCAA AAGCGTGGCT TTGTGCTTAT
51  GGAATTACTC ATGTCGTTCA CTCTAATTGC TTTGTTATTA GGGACTTTAG
151 GATTTTGGTA TCGGAAAATT TATACTGTAC AAAAGCAAAA AGAACGTATT
151 TATAACTTTT ATATCGAAGA AAGCCGAGCC TACAAGCAGC TCAGAACCCCT
201 GTTTAGCATG TCCTGTCTTT CATCTTACGA GGAGCTGGA TCATTATTTT
251 CTTTAACTCT TGATCGGGGT GTTTATCGAG ATCCTAAGCT GGCAGGTGCG
301 GTACGAGCTT CTCCTCATCA TGACACCAAG GATCAGAGAT TGGAACTTCG
201 TATTTGTAAT ATTAAGGATC AGTCTTACTT TGAAACACAG CGACTGCTCT
401 CCCACGTGAC CCATGTTGTA CTTTCCTTCC AGAGAAATCC TGATCCTGAA
451 AAACCTCCTG AAACAATTGC TTTAACTATA ACACGGGAAC CTAAGCATA
501 TCCTCCAAGG ACGTTAACAT ACCAATTTGC GGTGGGAAA TAA

```

The PSORT algorithm predicts a periplasmic location (0.915).

25 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 57A) or as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 57B) and FACS (Figure 57C) analyses.

These experiments show that cp7132 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 58

The following *C.pneumoniae* protein (PID 4376733) was expressed <SEQ ID 115; cp6733>:

```

1  MKTSIPWLV SSVLAFSCHL QSLANEELLS PDDSFNGNID SGTFTPKTSA
51  TTYSLTGDFV FYEPGKGTP LSDSCFKQTTD NLTFLGNGHS LTFGFIDAGT
101 HAGAAASTTA NKNLTFSGFS LLSFDSSPST TVTTGQGTLS SAGGVNLENI
151 RKLVVAGNFS TADGGAIKGA SFLLTGTS GD ALFSNNSSST KGGAIATTAG
201 ARIANNTGYV RFLSNIAS TS GGAIIDEGTS ILSNNKFLYF EGNAAKT TGG
251 AICNTKASGS PELIISNNKT LIFASNVAET SGGAIHAKKL ALSGGGFTEF
301 LRNNVSSATP KGGAISIDAS GELSLSAETG NITFVRNTLT TTGSTDTPKR
351 NAINIGSNGK FTELRAAKNH TIFYDPITS EGTSSDVLKI NNGSAGALNP
401 YQGTILFSGE TLTADLKV A DNLKSSFTQP VSLSGGKLLL QKGVLTLEST
451 FSQEAGSLLG MDSGTTLTST AGSITITNLG INVDSLGLKQ PVSLTAKGAS
501 NKVIVSGKLN LIDIEGNIYE SHMFSDQLF SLLKITVDAD VDTNVDISSL
551 IPVPAEDPNS EYGFQGWNV NWTDTATNT KEATATWTKT GFVPSPERKS
601 ALVCNTLWGV FTDIRSLQQL VEIGATGMEH KQGFVWSSMT NFLHKTGDEN
451 RKGFRHTSGG YVIGGSAHTP KDDLFTFAFC HLFARDKDCF IAHNNSRYSY
701 GTLFFKHSHT LQPNYLRLG RAKFSESAIE KFPREIPLAL DVQVSFSDSD
751 NRMETHYTS L PESEGSWSNE CIAGGIGLDL PFVLSNPHPL FKTFIPQMKV
801 EMVYVSQNSF FESSSDGRGF SIGRLNLST PVGAKFVQGD IGDSTYDLS

```

851 GFFVSDVYRN NPQSTATLVM SPDSWKIRGG NLSRQAFLLR GSNNYVYNSN
901 CELFGHYAME LRGSRRNYNV DVGTKLRF*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

```

5      1 ATGAAGACTT CGATTCCCTG GGTTTTAGTT TCCTCCGTGT TAGCTTTCTC
      51 ATGTCACCTA CAGTCACCTAG CTAACGAGGA ACTTTTATCA CCTGATGATA
     101 GCTTTAATGG AAATATCGAT TCAGGAACGT TTTACTCCAA AACTTCAGCC
     151 ACAACATATT CTCTAACAGG AGATGTCTTC TTTTACGAGC CTGGAAGAGG
     201 CACTCCCTTA TCTGACAGTT GTTTAAGCA AACCACGGAC AATCTTACCT
     251 TCTTGGGGAA CGGTCATAGC TTAACGTTTG GCTTTATAGA TGCTGGCACT
     301 CATGCAGGTG CTGCTGCATC TACAACAGCA AATAAGAATC TTACCTTCTC
     351 AGGGTTTTC TTTACTGAGT TTGATTCCTC TCCTAGCACA ACGGTTACTA
     401 CAGGTCAGGG AACGCTTTC TCAGCAGGAG GCGTAAATTT AGAAAATATT
     451 CGTAAACTTG TAGTTGCTGG GAATTTTCT ACTGCAGATG GTGGAGCTAT
     501 CAAAGGAGCG TCTTTCCTTT TAACTGGCAC TTCTGGAGAT GCTCTTTTAA
     551 GTAACAACTC TTCATCAACA AAGGGAGGAG CAATTGCTAC TACAGCAGGC
     601 GCTCGCATAG CAAATAACAC AGGTTATGTT AGATTCTCTAT CTAACATAGC
     651 GTCTACGTCA GGAGGCGCTA TCGATGATGA AGGCACGTCG ATACTATCGA
     701 ACAACAAATT TCTATATTTT GAAGGGAATG CAGCGAAAC TACTGGCGGT
     751 GCGATCTGCA ACACCAAGGC GAGTGGATCT CCTGAACCTGA TAATCTCTAA
     801 CAATAAGACT CTGATCTTTG CTTCAAACGT AGCAGAAACA AGCGGTGGCG
     851 CCATCCATGC TAAAAAGCTA GCCCTTTCCT CTGGAGGCTT TACAGAGTTT
     901 CTACGAAATA ATGTCTCATC AGCAACTCCT AAGGGGGGTG CTATCAGCAT
     951 CGATGCCCTCA GGAGAGCTCA GTCTTCTGTC AGAGACAGGA AACATTACCT
    1001 TTGTAAGAAA TACCCTTACA ACAACCGGAA GTACCGATAC TCCTAAACGT
    1051 AATGCGATCA ACATAGGAAG TAACGGGAAA TTCACGGAAT TACGGGCTGC
    1101 TAAAAATCAT ACAATTTTCT TCTATGATCC CATCACTTCA GAAGGAACCT
    1151 CATCAGACGT ATTGAAGATA AATAACGGCT CTGCGGGAGC TCTCAATCCA
    1201 TATCAAGGAA CGATTCTATT TTCTGGAGAA ACCCTAACAG CAGATGAACT
    1251 TAAAGTTGCT GACAATTTAA AATCTTCATT CACGCAGCCA GTCTCCCTAT
    1301 CCGGAGGAAA GTTATTGCTA CAAAAGGGAG TCACTTTAGA GAGCACGAGC
    1351 TTCTCTCAAG AGGCCGGTTC TCTCCTCGGC ATGGATTTCAG GAACGACATT
    1401 ATCAACTACA GCTGGGAGTA TTACAATCAC GAACCTAGGA ATCAATGTTG
    1451 ACTCCTTAGG TCTTAAGCAG CCCGTCAGCC TAACAGCAAA AGGTGCTTCA
    1501 AATAAAGTGA TCGTATCTGG GAAGCTCAAC CTGATTGATA TTGAAGGGAA
    1551 CATTTATGAA AGTCATATGT TCAGCCATGA CCAGCTCTTC TCTCTATTAA
    1601 AAATCAGCGT TGATGCTGAT GTTGATACTA ACGTTGACAT CAGCAGCCTT
    1651 ATCCCTGTTC CTGCTGAGGA TCCTAATTCA GAATACGGAT TCCAAGGACA
    1701 ATGGAATGTT AATTGGACTA CGGATACAGC TACAAATACA AAAGAGGCCA
    1751 CGGCAACTTG GACCAAAACA GGATTTGTTT CCAGCCCCGA AAGAAAATCT
    1801 GCGTTAGTAT GCAATACCCT ATGGGGAGTC TTTACTGACA TTCGCTCTCT
    1851 GCAACAGCTT GTAGAGATCG GCGCAACTGG TATGGAACAC AAACAAGGTT
    1901 TCTGGGTTTC TCCTATGACG AACTTCCTGC ATAAGACTGG AGATGAAAAT
    1951 CGCAAAGGCT TCCGTCATAC CTCTGGAGGC TACGTCATCG GTGGAAGTGC
    2001 TCACACTCCT AAAGACGACC TATTTACCTT TCGGTTCTGC CATCTCTTTG
    2051 CTAGAGACAA AGATTGTTTT ATCGCTCACA ACAACTCTAG AACCTACGGT
    2101 GGAACTTTAT TCTTCAAGCA CTCTCATACC CTACAACCCC AAAACTATTT
    2151 GAGATTAGGA AGAGCAAAGT TTTCTGAATC AGCTATAGAA AAATTCCTCA
    2201 GGGAAATTCC CCTAGCCTTG GATGTCCAAG TTTCGTTTCAG CCATTTCAGAC
    2251 AACCGTATGG AAACGCACTA TACCTCATTG CCAGAATCCG AAGGTTCTTG
    2301 GAGCAACGAG TGTATAGCTG GTGGTATCGG CCTAGACCTT CCTTTTGTTC
    2351 TTTCCAACCC ACATCCTCTT TTCAAGACCT TCATTCCACA GATGAAAGTC
    2401 GAAATGGTTT ATGTATCACA AAATAGCTTC TTCGAAAGCT CTAGTGATGG
    2451 CCGTGGTTTT AGTATTGGAA GGCTGCTTAA CCTCTCGATT CCTGTGGGTG
    2501 CGAAATTCGT GCAGGGGGAT ATCGGAGATT CCTACACCTA TGATCTCTCA
    2551 GGATTCCTTG TTTCCGATGT CTATCGTAAC AATCCCAAT CTACAGCGAC
    2601 TCTTGTGATG AGCCAGACT CTTGGAAGAT TCGCGGTGGC AATCTTCAA
    2651 GACAGGCATT TTTACTGAGG GGTAGCAACA ACTACGTCTA CAACTCCAAAT
    2701 TGTGAGCTCT TCGGACATTA CGCTATGGAA CTCCGTGGAT CTTCAAGGAA
    2751 CTACAATGTA GATGTTGGTA CCAAACCTCCG ATTCTAG

```

The PSORT algorithm predicts an outer membrane location (0.924).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 58A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 58B) and for FACS (Figure 58C) analyses. A GST-fusion protein was also expressed.

The cp6733 protein was also identified in the 2D-PAGE experiment (Cpn0451).

- 5 These experiments show that cp6733 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 59

The following *C.pneumoniae* protein (PID 4376814) was expressed <SEQ ID 117; cp6814>:

```

10      1  MHDALLSILA IQELDIKMIR LMRVKKEHQK ELAKVQSLKS DIRRKVQEKE
      51  LEMENLKTQI RDGENRIQEI SEQINKLENQ QAAVKKMEDEF NALTQEMTTA
     101  NKERRSLEHQ LSDLMDKQAG GEDLIVSLKE SLASTENSSS VIEKEIFESTI
     151  KKINEEGKAL LEQRTTELKHA TNPELLSIYE RLLNNKKDRV VVPIENRVCS
     201  GCHIVLTPQH ENLVRKKDRL IFCEHCSRIL YWQESQVNAQ ENSTAKRRRR
     251  RAAV*
```

- 15 The cp6814 nucleotide sequence <SEQ ID 118> is:

```

      1  ATGCATGACG CACTTCTAAG CATTTTGGCT ATTCAAGAGC TTGATATTAA
      51  AATGATTTCG CTTATGCGCG TAAAGAAAGA ACATCAGAAA GAATTGGCTA
     101  AAGTCCAATC TTTAAAAAGT GATATTTCGT GAAAAGTTCA GAAAAAGAA
     151  CTCGAAATGG AGAATTTGAA AACTCAAATT CGAGATGGAG AGAATCGCAT
     201  CCAAGAGATT TCTGAACAAA TCAATAAATT AGAAAATCAG CAAGCTGCTG
     251  TAAAAAAAT GGATGAGTTT AACGCTCTTA CCCAAGAAAT GACTACAGCA
     301  AACAAAGAAC GTCGCTCTTT AGAGCACCAG CTTAGCGATC TCATGGATAA
     351  GCAAGCTGGA GGCGAAGACC TTATTGTCTC TCTAAAAGAA AGCTTAGCTT
     401  CTACAGAAAA TAGTAGCAGT GTCATTGAAA AAGAAATTTT TGAAAGCATC
     451  AAAAAGATTA ATGAAGAAGG CAAAGCTTTG CTTGAACAAC GGACAGAGTT
     501  AAAGCATGCG ACGAATCCCG AACTACTCAG CATCTATGAG CGTCTATTAA
     551  ACAATAAAAA AGATCGCGTT GTTGTTCCTA TTGAAAATCG TGTCTGCAGT
     601  GGTGTGCATA TTGTTCTAAC TCCTCAACAC GAAAATCTTG TAAGAAAGAA
     651  AGACCGACTC ATTTTTCGCG AACATTGCTC TCGAATTCTC TATTGGCAAG
     701  AATCCCAAGT CAATGCTCAG GAAAATTCCA CAGCAAAACG TCGTCGTCGT
     751  CGCGCAGCTG TATAA
```

The PSORT algorithm predicts an inner membrane location (0.070).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 59A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 59B) and FACS (Figure 59C) analyses.

These experiments show that cp6814 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 60

The following *C.pneumoniae* protein (PID 4376830) was expressed <SEQ ID 119; cp6830>:

```

40      1  MKWLPTAVF AAVLPALTAF GDPASVEIST SHTGSGDPTS DAALTGFTQS
      51  STETDGTYYT IVGDITFSTF TNIPVPVVTP DANDSSSNSS KGGSSSSGAT
     101  SLIRSSNLHS DFDFTKDSVL DLYHLFFPSA SNTLNPALLS SSSSSGSSSS
     151  SSSSSGSAS AVVAADPKGG AAFYSNEANG TLFTTDSGN PGSLTLQNLK
     201  MTGDGAIIYS KGPLVFTGLK NLTFTGNESQ KSGGAAYTEG ALTTQAIVEA
     45  251  VTFTGNTSAG QGGAIYVKEA TLFNALDSLK FEKNTSGQAG GGIYTESTLT
     301  ISNITKSIEF ISNKASVFAP APEPTSPAPS SLINSTTIDT STLQTRAASA
     351  TPAVAPVAHV TPTPISTQET AGNGGAIYAK QGISISTFKD LTFKSNASAV
```

5 401 DATLTVDSST IGESGGAIFA ADSIQIQQCT GTTLFSGNTA NKSGGGIYAV
 451 GQVTLLEDIAN LKMTNNTCKG EGGAIYTKKA LTINNGAILT TFGNTSTNDN
 501 GGAIFAVGGI TSLDLVEVRF SKNKTGNYS A PITKAASNTA PVSSTSTAA
 551 SPAVPAAAAA PVTNAAKGGA LYSTEGTLVS GITSILSFEN NECQNQGGGA
 601 YVTKTFQCS D SHRLQFTSNK AADEGGGLYC GDDVTLTNLT GKTLPQENSS
 651 EKHGGGLSLA SGKSLTMTSL ESFCLNANTA KENGGGANVP ENIVLTFITYT
 701 PTPNEPAPVQ QPVYGEALVT GNTATKSGGG IYTKNAAFSN LSSVTFDQNT
 751 SSENGGALLT QKAADKTDCS FTYITNVNIT NNTATGNGGG IAGGKAHFDR
 801 IDNLTVQSNQ AKKGGGVYLE DALILEKVIT GSVSQNTATE SGGGIYAKDI
 10 851 QLQALPGSFT ITDNKVETSL TTSTNLYGGG IYSSGAVTLT NISGTFGITG
 901 NSVINTATSQ DADIQGGGIY AT'TSLINQC NTPILFSNNS AATKKTSTTK
 951 QIAGGAIFSA AVTIENNSQP IIFLNNSAKS EATTAATAGN KDCSGGAIAA
 1001 NSVTLTNPE ITFKGNYAET GGAIGCIDLT NGSPPRKYSI ADNGSVLFDQ
 1051 NSALNRGGAI YGETIDISRT GATFIGNSSK HDGSAICCST ALTLAPNSQL
 15 1101 IFENNKVTET TATTKASINN LGAAIYGNNE TSDVTISLSA ENGSIFFKNN
 1151 LCTATNKYCS IAGNVKFTAI EASAGKASF YDAVNVSTKE TNAQELKLINE
 1201 KATSTGTILF SGELHENKSY IPQKVTFAHG NLILGKNAEL SVVSFTQSPG
 1251 TTTITMGPFSV LSNHSKEAGG IAINNVILDF SEIVPTKDNA TVAPPTLKLIV
 20 1301 SRTNADSKDK IDITGTVTL DPNGNLYQNS YLGEDRDITL FNIDNSASGA
 1351 VTATNVTLQG NLGAKKGYLG TWNLDPNSSG SKIILKWTFD KYLRWPYIPR
 1401 DNHFYINSIW GAQNSLVTVK QGILGNMLNN ARFEDPAPNN FWASAIGSHL
 1451 RKEVSRNSDS FTYHGRGYTA AVDAKPROEF ILGAAFSQVF GHAESEYHLD
 1501 NYKHKSGSHS TQASLYAGNI FYFPAIRSRP ILFQGVATYG YMQHDTTTYT
 25 1551 PSIEEKNMAN WDSIAWLFDL RFSVDLKEPQ PHSTARLTIFY TEABYTRIRQ
 1601 EKFTELDYPD RSFSACSYGN LAIPTGFSVD GALAWREIIL YNKVSAAYLP
 1651 VILRNPKAT YEVLSTKEG NVVNVLPTRN AARAEVSSQI YLGSYWTLYG
 1701 TYTIDASMNT LVQMANGGIR FVF*

A predicted signal peptide is highlighted.

The cp6830 nucleotide sequence <SEQ ID 120> is:

30 1 ATGAAGTGGC TACCAGCTAC AGCTGTTTTT GCTGCCGTAC TCCCCGCACT
 51 AACAGCCTTC GGAGATCCCG CGTCTGTGTA AATAAGTACC AGCCATACAG
 101 GATCCGGGGA TCCTACAAGC GACGCTGCCT TAACAGGATT TACACAAAGT
 151 TCCACAGAAA CTGACGGTAC TACCTATACC ATTGTCCGTG ATATCACCTT
 201 CTCTACTTTT ACGAATATTC CTGTTCCCGT AGTAACTCCA GACGCCAACG
 35 251 ATAGTTCAG CAATAGCTCT AAAGGAGGAA GTAGCAGTAG TGGAGCTACA
 301 TCTCTAATCC GATCCTCAAA CCTACACTCC GATTTTGATT TTACAAAAGA
 351 TAGCGTGTTA GACCTCTATC ACCTTTTCTT TCCTTCAGCT TCAAACTATC
 401 TCAATCCTGC ACTCCTTTCT TCCAGTAGCA GCGGTGGATC CTCGAGCAGC
 451 AGTAGCTCCT CATCATCTGG AAGTGCACTC GCTGTTGTGTG CTGCGGACCC
 40 501 AAAAGGAGGC GCTGCCCTTT ATAGTAACGA GGCTAACGGA ACTTTAACCT
 551 TCACTACAGA CTCTGGAAAT CCCGGCTCCC TGACTCTTCA GAATCTTAAA
 601 ATGACCGGAG ATGGAGCCGC CATCTACTCG AAGGGTCCCTC TAGTATTTAC
 651 TGGTTTAAAA AATCTAACCT TTACAGGAAA TGAATCTCAG AAATCTGGAG
 701 GTGCTGCCTA TACTGAAGGC GCACTCACA CACAAGCAAT CGTTGAAGCC
 45 751 GTAACCTTTA CTGGCAACAC CTCGGCAGGG CAAGGAGCGC CTATCTATGT
 801 TAAAGAAGCT ACCCTATTCA ATGCTCTAGA CAGCCTCAAA TTTGAAAAAA
 851 ACACTTCTGG GCAAGCTGGT GGTGGAATCT ATACAGAGTC TACGCTCACA
 901 ATCTCGAACA TCACAAAATC TATTGAATTT ATCTCTAATA AAGCTTCTGT
 951 CCCTGCCCCC GCTCCTGAGC CCACCTCTCC GGCTCCAAGT AGCTTAATAA
 50 1001 ATTCTACAAC GATCGATACC TCGACTCTCC AAACCCGAGC AGCATCCGCA
 1051 ACTCCAGCAG TGGCTCCTGT TGCTGCCGTA ACTCCAACAC CAATCTCTAC
 1101 TCAAGAGACC GCAGGAAATG GAGGCGCTAT CTATGCTAAA CAAGGTATTT
 1151 CGATATCCAC GTTTAAAGAT CTGACCTTCA AGTCTAACTC TGCATCGGTA
 1201 GATGCCACCC TTAAGTCTGA TTCTAGCACT ATTGGAGAAT CTGGAGGTGC
 55 1251 TATCTTTGCA GCAGACTCTA TACAAATCCA ACAGTGCACG GGAACCACTT
 1301 TATTCTAGTG CAATACTGCC AATAAGTCTG GTGGGGGTAT TTACGCTGTA
 1351 GGACAAGTCA CCCTAGAAGA TATAGCGAAT CTGAAGATGA CCAACACAC
 1401 CTGTAAGGT GAAGGTGGAG CCATCTACAC TAAAAAGGCT TTAACATCA
 1451 ACAACGGTGC CATTCTCACT ACATTTTCTG GAAATACATC GACAGATAAT
 60 1501 GGTGGGGCTA TTTTGTCTGT AGGTGGCATC ACTCTCTCTG ATCTTGTAGA
 1551 AGTCCGCTTT AGTAAAAATA AGACCGGAAA TTATTCGCTC CTATTACCA
 1601 AAGCGGCTAG CAACACAGCT CCGTAGTTT CTAGCTCTAC AACTGCTGCA
 1651 TCTCCTGCGG TCCCTGCTGC CGCTGCAGCA CCTGTTACAA ACGCAGCAAA
 1701 AGGAGGGGCT TTATATAGTA CAGAAGGACT GACTGTATCT GGAATCACAT
 65 1751 CGATATTGTC GTTTGAAAAC AACGAATGCC AGAATCAAGG AGGTGGGGCT

	1801	TACGTTACTA	AAACCTTCCA	GTGTTCCGAT	TCTCATCGCC	TCCAGTTTAC
	1851	TAGTAATAAA	GCAGCAGATG	AAGGCGGGGG	CCTGTATTGT	GGTGACGATG
	1901	TCACGCTAAC	GAACCTGACA	GGGAAAACAC	TATTTCAAGA	GAATAGCAGT
5	1951	GAGAAACATG	GAGGTGGGCT	CTCTCTCGCC	TCAGGAAAAT	CTCTGACTAT
	2001	GACATCGTTA	GAGAGCTTCT	GCTTAAATGC	AAATACAGCA	AAGGAAAACG
	2051	GAGGCGGTGC	GAATGTCCCT	GAAAAATATT	TACTCACCTT	CACCTATACT
	2101	CCCCTCCAA	ATGAACCTGC	GCCTGTGCAG	CAGCCCGTGT	ATGGAGAAGC
	2151	TCTTGTACT	GGAAATACAG	CCACAAAAG	TGGTGGGGGC	ATTTACACGA
10	2201	AAAAATGCGG	CTTCTCAAAT	TTATCTTCTG	TAAC'TTTGA	TCAAAATACC
	2251	TCTTCAGAAA	ATGGTGGTGC	CTTACTTACC	CAAAAAGCTG	CAGATAAAAC
	2301	GGACTGTTCT	TTACCTTATA	TTACAAATGT	CAATATCACC	AACAATACAG
	2351	CTACAGGAAA	TGGTGGGGGC	ATTGCTGGGG	GAAAAGCACA	TTTCGATCGC
	2401	ATTGATAATC	TTACAGTCCA	AAGCAACCAA	GCAAAGAAAG	GTGGTGGGGT
15	2451	TTATCTTGAA	GATGCCCTCA	TCCTGGAAAA	GGTTATTACA	GGTTC'TGTCT
	2501	CACAAAATAC	AGCTACAGAA	AGTGGTGGGG	GTATCTACGC	TAAGGATATT
	2551	CAACTACAAG	CTCTACCTGG	AAGCTTCACA	ATTACCGATA	ATAAAGTCGA
	2601	AACTAGTCTT	ACTACTAGCA	CTAATTTATA	TGGTGGGGGC	ATCTATTCCA
	2651	GTGGAGCTGT	CACGCTAACC	AATATATCTG	GAACCTTTGG	CATTACAGGA
20	2701	AACTCTGTTA	TCAATACAGC	GACATCCAG	GATGCAGATA	TACAAGGTGG
	2751	GGGCATTTAT	GCAACACAGT	CTCTCTCAAT	AAATCAATGT	AATACACCCA
	2801	TTCTATTTAG	CAACAACCTC	GCTGCCACTA	AAAAACATC	AACAACAAAG
	2851	CAAAATGCTG	GTGGGGCTAT	CTTCTCCGCT	GCAGTAACTA	TCGAGAATAA
	2901	CTCTCAGCCC	ATTATTTTCT	TAAATAATTC	CGCAAAGTCG	GAAGCAACTA
25	2951	CAGCAGCAAC	TGCAGGAAAT	AAAGATAGCT	GTGGAGGAGC	CATTGCAGCT
	3001	AACTCTGTTA	CTTTAACAAA	TAACCTTGAA	ATAACCTTTA	AAGGAAATTA
	3051	TGCAGAAACT	GGAGGAGCGA	TTGGCTGTAT	TGATCTTACT	AATGGCTCAC
	3101	CTCCCGGTAA	AGTCTCTATT	GCAGACAACG	GTTCTGTCTT	TTTTCAGAC
	3151	AACTCTGCGT	TAAATCGCGG	AGGCGCTATC	TATGGAGAGA	CTATCGATAT
30	3201	CTCCAGGACA	GGTGCGACTT	TCATCGGTAA	CTCTTCAAAA	CATGATGGAA
	3251	GTGCAATTTG	CTGTTCAACA	GCCCTAAACT	TTGCCCAAAA	CTCCCAACTT
	3301	ATCTTTGAAA	ACAATAAGGT	TACGGAACCC	ACAGCCACTA	CAAAAGCTTC
	3351	CATAAATAAT	TTAGGAGCTG	CAATTTATGG	AAATAATGAG	ACTAGTGACG
	3401	TCACTATCTC	TTTATCAGCT	GAGAATGGAA	GTATTTTCTT	TAAAAACAAT
35	3451	CTATGCACAG	CAACAAACAA	ATACTGCAGT	ATTGCTGGAA	ACGTAAAATT
	3501	TACAGCAATA	GAAGCTTCAG	CAGGGAAAGC	TATATCTTTC	TATGATGCAG
	3551	TTAACGTTTC	CACCAAAGAA	ACAAATGCTC	AAGAGCTAAA	ATTAATAGAA
	3601	AAAGCGACAA	GTACAGGAAC	GATTTCTATTT	TCTGGGGAAC	TTCCAGAAAA
	3651	TAAATCCTAT	ATTCCACAGA	AAGTCACTTT	CGCACATGGG	AATCTCATTC
40	3701	TAGGTAAAAA	TGCAGAACTT	AGCGTAGTTT	CCTTTACCCA	ATCTCCAGGC
	3751	ACCACAATCA	CTATGGGGCC	AGGATCGGTT	CTTTCCAACC	ATAGCAAGAA
	3801	AGCAGGAGGA	ATCGCTATAA	ACAAATGTCAT	CATTGATTTT	AGTGAAAATCG
	3851	TTCTACTATA	AGATAATGCA	ACAGTAGCTC	CACCCACTCT	TAAATTAGTA
	3901	TCGAGAACTA	ATGCAGATAG	TAAAGATAAG	ATTGATATTA	CAGGAACCTGT
45	3951	GACTCTTCTA	GATCTTAATG	GCAACTTATA	TCAAAATTCT	TATCTTGGTG
	4001	AAGACCGCGA	TATCACTCTT	TTCAATATAG	ACAA'TTCTGC	AAGTGGGGCA
	4051	GTTACAGCCA	CGAATGTAC	CCTTCAAGGG	AA'TTTAGGAG	CTAAAAAAGG
	4101	ATATTTAGGA	ACCTGGAATT	TGGATCCAAA	TTCTCGGGT	TCAAAAATTA
	4151	TTCTAAAATG	GACCTTTGAC	AAATACCTGC	GCTGGCCCTA	CATCCCTAGA
50	4201	GACAACCACT	TCTACATCAA	CTCTATTTGG	GGAGCACAAA	ACTCTTTAGT
	4251	GACTGTGAAA	CAAGGGATCT	TAGGGAACAT	GTTGAACAA	GCAAGGTTTG
	4301	AAGATCCTGC	TTTCAACAAC	TTCTGGGCTT	CGGCTATAGG	ATCTTTCTTT
	4351	AGGAAAGAAG	TATCTCGAAA	TTCTGACTCA	TTCACCTATC	ATGGCAGAGG
	4401	CTATACCGCT	GCTGTGGATG	CCAAACCTCG	CCAGAATTTT	ATTTTAGGAG
55	4451	CTGCCTTCAG	TCAGGTTTTT	GGTCACGCCG	AGTCTGAATA	TCACCTTGAC
	4501	AACTATAAGC	ATAAAGGCTC	AGGTCACTCT	ACACAAGCAT	CTCTTTATGC
	4551	TGGCAATATC	TTCTATTTTC	CTGCGATACG	GTCTCGGCCT	ATCTATTTCC
	4601	AAGGTGTGGC	GACCTATGGT	TATATGCAAC	ATGACACCAC	AACTACTAT
	4651	CCTTCTATTG	AAGAAAAAAA	TATGGCAAAC	TGGGATAGCA	TTGCTTGGTT
60	4701	ATTTGATCTG	CGTTTCAGTG	TGGATCTTAA	AGAACCTCAA	CCTCACTCTA
	4751	CAGCAAGGCT	TACCTTCTAT	ACAGAAGCTG	AGTATACCAG	AATTCGCCAG
	4801	GAGAAATTCA	CAGAGCTAGA	CTATGATCCT	AGATCTTTCT	CTGCATGCTC
	4851	TTATGGAAAC	TTAGCAAATC	CTACTGGATT	CTCTGTAGAC	GGAGCATTAG
	4901	CTTGGCGTGA	GATTATTCTA	TATAATAAAG	TATCAGCTGC	GTACCTCCCT
65	4951	GTGATTCTCA	GGAATAATCC	AAAAGCGACC	TATGAAGTTC	TCCTACAAA
	5001	AGAAAAGGGC	AACGTAGTCA	ACGTTCTCCC	TACAAGAAAC	GCAGCTCGTG
	5051	CAGAGGTGAG	CTCTCAAATT	TATCTTGGAA	GTTACTGGAC	ACTCTACGGC
	5101	ACGTATACTA	TTGATGCTTC	AATGAATACT	TTAGTGCAAA	TGGCCAACGG
	5151	AGGGATCCGG	TTTGTATTCT	AG		

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 60A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 60B) and FACS (Figure 60C) analyses.

- 5 The cp6830 protein was also identified in the 2D-PAGE experiment (Cpn0540) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6830 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 61

- 10 The following *C.pneumoniae* protein (PID 4376854) was expressed <SEQ ID 121; cp6854>:

```

1  MSIAIAREQY AAILDMHPKP SIAMFSSEQA RTSWEKROAH PYLYRLLEII
51  WGVVKFLLGL IFFIPLGLFW VLQKICQNF ILLGAGGWIFR PICRDSNLLR
101 QAYAARLFSA SFQDHVSSVR RVCLQYDEVF IDGLELRLPN AKPDRWMLIS
151 NGNSDCLEYR TVLQGEKDWI FRIAESQSN ILIFNYPGVM KSQLNITRNN
15  201 VVKSQACVR YLRDEPAGPQ ARQIVAYGYS LGASVQAEAL SKEIADGSDS
251 VRWFVVKDRG ARSTGAVAKQ FIGSLGVWLA NLTHWNINSE KRSKDLHCPE
301 LFIYKDSQG NLIGDGLFKK ETCFAAPFLD PKNLEECSGK KIPVAQTGLR
351 HDHILSDDVI KEVAGHIQRH FDN*
```

The cp6854 nucleotide sequence <SEQ ID 122> is:

```

20 1  ATGTCAATAG CTATTGCAAG GGAACAATAC GCAGCTATAT TGGATATGCA
51  TCCTAAACCT TCGATCGCCA TGTTCCTTTC GGAGCAGGCG AGAATTCTTT
101 GGGAGAAACG ACAGGCTCAT CCTTACCTTT ATCGTCTTCT TGAGATCATA
151 TGGGGTGTG TGAAATTTCT TCTCGGCTTA ATCTTCTTTA TTCCTTGGG
201 TCTTTTCTGG GTCCTTCAGA AGATATGTCA GAATTTTATT CTCTTTGGTG
25 251 CAGGAGGGTG GATTTTTAGA CCCATATGCA GGGACTCTAA TTTATTGCGA
301 CAAGCTTACG CCGCGCTCT TTTCTCCGCT TCATTCCAAG ATCATGTCTC
351 CTCTGTGCGA AGGTTTGCT TACAGTATGA CGAGGTCTTT ATTGACGGAT
401 TGGAGTTACG TCTTCCCAAT GCTAAGCCAG ATCGATGGAT GTPAATCTCC
451 AATGGAAACT CCGATTGCTT AGAGTATAGG ACAGTGCTGC AAGGGGAAAA
30 501 GGACTGGATA TTCGTTATTG CTGAAGAGTC TCAATCCAAC ATTTTAATCT
551 TCAATTACCC AGGAGTCATG AAGAGCCAAG GGAATATAAC AAGAAACAAT
601 GTAGTCAAAT CTTATCAAGC ATGCGTACGC TATCTTAGAG ATGAACCCGC
651 AGGACCTCAG GCGCGTCAAA TCGTTGCTTA TGGCTATTCT TTAGGAGCTA
701 GTGTTCAAGC CGAAGCATTA AGTAAAGAGA TCGCAGACGG AAGTGATAGC
35 751 GTCCGTGGT TGTGCTTAA AGATCGAGGA GCTCGCTCTA CAGGAGCCGT
801 TGCTAAACAG TTTATTGGAA GTCTAGGAGT TTGGCTGGCG AATCTTACCC
851 ATTGGAATAT TAATCTGAA AAGAGAAGCA AGGACTTGCA TTGCCAGAA
901 CTCTTTATTT ATGGCAAGGA TTCCCAAGGT AATCTTATCG GGGATGGATT
951 GTTCAAAAAA GAGACGTGCT TCGCAGCACC ATTTTATAGT CCTAAAAACT
40 1001 TGAAGAGTG TTCAGGGAAG AAAATCCCTG TAGCTCAGAC CGTCTAAGA
1051 CACGATCATA TCCTTCCGA TGATGTGATT AAAGAAGTTG CAGGTCATAT
1101 TCAAAGACAT TTCGATAATT A
```

The PSORT algorithm predicts an inner membrane location (0.461).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 61A.

- 45 The recombinant protein was used to immunise mice, whose sera were used in Western blot (Figure 61B) and FACS (Figure 61C) analyses. A his-tagged protein was also expressed.

These experiments show that cp6854 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 62

The following *C.pneumoniae* protein (PID 4377101) was expressed <SEQ ID 123; cp7101>:

1	MYSCYSKGIS	HNVLHPMSR	LDIFVFDLSI	ANQDQNLLEE	IFCSEDTVLF
5	51	KAYRTTALQS	PLAAKNLNIA	RKVANYILAD	NGEIDTVKLV
	101	YPLGPHRHNE	AQDREHLLKM	LKALKENPKL	KESIKTLFVP
	151	HTLALNPQTI	LSTIHVRQAA	LTALFTYLRO	DVGSCEFATAP
	201	RFLKDLNDLI	SSGKLSRIVN	QREIAVPINL	SGCIGELFKP
	251	LVKLSSSPGL	KKAFSAANLI	ETLGDSEAQI	QQLLSHQYLM
	301	TANDIIKSTL	LHYYQLQEST	VRAIFPKEGL	FSKEQVAFST
10	351	RVYHYLHAYE	EAKSAFIHDT	QNPLLKAWAY	TLATLADASQ
	401	LGWKSSEDPHS	LVSLVTHFVE	EEVENIRILV	QQCEQTYHEA
	451	MRNPLNNQDS	QILTMHMRF	RQELNKALYE	WDSAQEKAKK
	501	FYTKQIPLYF	RSSYDAFIQE	FAHLYANAPA	GFRILFTHGR
	551	SINEFIRFLS	EFFTSTESL	LGKHAVINLE	KETSRLVHNI
15	601	EALLTRILEA	YQLPVPPSIL	NHLDQLSQTP	WVYVSGGTVD
	651	EPLTLTEKHP	ENPHELAIFY	ADALKDLPTG	IKSYLEEGSH
	701	FSIIAGSPLE	REAWDNDWYS	YTWLRDVWVK	QHDFLQDTI
	751	ENFCNKYALQ	HVVHDFHDFC	SDHSLTLPEL	YDKGSRFLSS
	801	IYIRRLLYLM	VREVPYVSEQ	QLPEVLNVNS	SYLGISSRIT
20	851	TIPKMTLLSS	ADLRHIYKGL	LMQSYQKIYT	EEDTYLRLFT
	901	APLLFADSNW	PSIYFGFILN	PGTTEIDLWK	FNAYAGLQGP
	951	SRPWTLYANP	IDYGMPPPPG	YRSRLPKEFF	*

The cp7101 nucleotide sequence <SEQ ID 124> is:

1	ATGTATTCGT	GTACAGCAA	AGGAATATCC	CATAACTATC	TTCTACATCC
25	51	TATGTCACGT	TTGGATATTT	TTGTTTTCGA	TTCTCTGATC
	101	ATCAAAATCT	TCTTGAGGAA	ATTTTCTGTT	CTGAAGACAC
	151	AAAGCCTACC	GTACTACGGC	TCTACAATCC	CCTCTAGCTG
	201	AAATATCGCC	CGTAAAGTCG	CAAATTTATAT	CTTAGCTGAC
	251	TCGATACAGT	AAAGCTTGTC	GAAGCCATTC	ACCATCTCTC
30	301	TATCCTTTAG	GGCCTCATCG	CCATAATGAA	GCTCAAGATC
	351	CCTTAAATAG	CTAAAAGCTC	TAAAGGAAAA	TCCTAAATTA
	401	TCAAAACCTC	CTTGTGCCCT	TCATACTCTA	CAATCCAAAA
	451	CATACACTAG	CATTGAATCC	ACAGACAATT	CTCTCTACGA
	501	TCAAGCAGCA	CTCACAGCGC	TCTTCACCTA	CCTTCGGCAA
35	551	CCTGTTTTCG	TACGGCTCCT	GCCATTCTCA	TTCACCAAGA
	601	CGATTCTCTA	AAGATCTCAA	TGATCTCATT	AGCAGTGGCA
	651	AATCGTAAAC	CAAAGGGAAA	TTGCGGTTCC	TATAAACCTT
	701	TTGGAGAGCT	ATTCAAGCCT	TAAAGGATTC	TAGATCTTTA
	751	CTGGTTAAGC	TCTCCTCATC	TCCAGGACTC	AAAAAAGCCT
40	801	CAATCTTATT	GAAACTCTTG	GGGATTCTGA	AGCACAATTC
	851	TCTCGCATCA	ATATTTGATG	CAAAAACCTAC	AAAATGTCCA
	901	ACTGCTAACG	ACATTTATCA	ATCGACACTT	CTGCACTACT
	951	AGAAAGTACT	GTACGAGCTA	TTTTCTTCAA	AGAAGGGTTG
	1001	AACAAGTGGC	ATTTCTCGACG	CAACACCCCA	GAGAGCTCTC
45	1051	CGGGTATACC	ACTACTTACA	TGCCTATGAA	GAAGCAAAAT
	1101	CCATGACACT	CAAAATCCCT	TACTGAAAGC	CTGGGAGTAT
	1151	CTCTTGCGGA	TGCTAGCCAA	CCTACCATCT	CAAACCATAT
	1201	TTAGGATGGA	AAAGTGAAGA	CCCTCACAGT	CTTGTATCTC
	1251	CTTTGTTGAA	GAGGAAGTAG	AAAACATCCG	AATTTTAGTC
50	1301	AACAGACCTA	TCACGAAGCA	CGCTCCCAAC	TAGAATATAT
	1351	ATGCGCAACC	CACTAAATAA	TCAAGACAGT	CAGATTTTGA
	1401	CATGCGCTTC	CGTCAAGAAC	TCAATAAAGC	TCTTTATGAG
	1451	CTCAAGAAAA	GGCAAGAAAA	TTTCTACATC	TTCTTGAATT
	1501	TTCTATACAA	AGCAAATTC	CTTATACTTT	CGTAGTTCTT
55	1551	CATTCAAGAA	TTTGCTCATC	TCTATGCTAA	TGCTCCCGCT
	1601	TTCTTTTTCAC	GCATGGACGC	ACCCATCCGA	ACACATGGTC
	1651	TCGATTAAATG	AAATTATACG	TTTTCTTTCT	GAATTCTTCA
	1701	GTCAGAACTT	CTGGGGAAC	ATGCGGTGAT	CAATTTAGAG
	1751	CTCGGCTCGT	CCACAACATC	ACTGCCATGC	TACACACGGA
60	1801	GAAGCTCTCC	TTACAAGAAAT	TTTAGAAGCC	TATCAGCTTC
	1851	CTCCATCTTA	AACCACTTAG	ATCAGCTGTC	ACAAACTCCC
	1901	TTTCTGGAGG	AACAGTGGAC	ACTCTTCTTT	TGGATTATTT
	1951	GAACCTCTGA	CACTTACAGA	AAAGCATCCT	GAAATCTCTC
	2001	AGCTTCTTAC	GCAGACGCCC	TTAAAGATCT	CCCTACAGGA

-103-

2051 ATCTAGAAGA AGGATCCCAC TCTCTACTTA GCTCATCACC CACCCACGTT
 2101 TTCTCTATAA TCGCAGGATC TCCTTTATTT CGGGAAGCTT GGGATAATGA
 2151 TTGGTACAGC TATACCTGGC TTCGTGATGT CTGGGTGAAA CAACACCAAG
 2201 ATTTCCTTCA AGATACTATA TTACCTCAGC TAAGTATCTA TGCTTTCTATA
 2251 GAGAATTTT GTAAACAAATA TGCTTTGCAA CATGTAGTTC ATGACTTTCA
 2301 TGATTTCTGC TCCGACCACT CCTTGACTCT TCCGGAGCTC TATGACAAAG
 2351 GATCGCGTTT TCTAAGCTCC TTATTCACCA AAGATAAGAC CGTAGCTCTT
 2401 ATCTATATAC GCCGTCTTCT CTACCTTATG GTCCGTGAAG TCCCTTATGT
 2451 TTCAGAACAA CAGCTTCCAG AAGTCTTAGA TAACGTCTCT TCATATCTCG
 2501 GGATTTCCCT TCGTATTACC TATGAGAAAT TCCGCTCCCT GATAGAGGAA
 2551 ACCATCCCTA AAATGACCTT ACTCTCTCA GCAGACCTGA GGCATATCTA
 2601 TAAAGGTCTC CTATGCAAA GTTATCAAAA GATCTACACC GAAGAAGATA
 2651 CGTACCTCCG CCTCACCACG GCAATGAGGC ATCATAATCT TGCCTATCCC
 2701 GCTCCTTTGC TCTTTGCAGA CAGTAACTGG CCTTCTATTT ATTTTGGATT
 2751 CATCCTAAAT CCAGGAACCA CAGAGATCGA TCTTTGGAAA TTAACTATG
 2801 CAGGGCTGCA AGGACAGCCT CTGACAATA TCCAGGAGCT GTTCGCAACG
 2851 TCAAGACCTT GGACCTCTA TGCAATCCT ATAGATTATG GCATGCCACC
 2901 GCCTCCAGGC TACCGCAGCC GCCTCCCTAA AGAATTTTTC TAG

The PSORT algorithm predicts a cytoplasmic location (0.206).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 62A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 62B) and FACS (Figure 62C) analyses.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 25 These experiments show that cp7101 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 63

The following *C.pneumoniae* protein (PID 4377107) was expressed <SEQ ID 125; cp7107>:

1 MSIVRNSALP LPCLSRSETF KVRSHMKFM KVLTPWIYRK DLWVTAPLLT
 51 AIPGSFAHTL VDIAGEPRHA AQATGVSGDG KIVIGMKVPD DPFAITVGFQ
 101 YIDGHLQPLE AVRPOCSVYP NGITPDGTVI VGTNYAIGMG SVAVKRWNGK
 151 VSELPMLPDT LDSVASAVSA DGRVIGGNRN INLGASAVK WEDDVITQLP
 201 SLPDAMNACV NGISSDGSII VGTMDVSWR NTAQVQWIGDQ LSVIGTLGGT
 251 TSVASAISTD GTVIVGGSSEN ADSQTHAYAY KNGVMSDIGT LGGFYSLAHA
 301 VSSDGSVIVG VSTNSEHRYH AFQYADGQMV DLGTLGGPES YAQGVSGDGK
 351 VIVGRAQVPS GDWHAFPCPF QAPSPAPVHG GSTVTVTSQNP RGMVDINATY
 401 SSLKNSQQQL QRLLIQHSK VESVSSGAPS FTSVKGAISK QSPAVQNDVQ
 451 KGTFLSYRSQ VHGNVQNQQ L TGAFMDWKL ASAPKCGFKV ALHYGSQDAL
 501 VERAALPYTE QGLGSSVLSG FGGQVQGRYD FNLGETVVLO PFMGIQVLHL
 551 SREGYSEKNV RFPVSYDSVA YSAATSFMDA HVFASLSPKM STAATLGVLR
 601 DLNSHIDEFK GSVSAMGNFV LENSTVSVLR PFASLAMYD VRQQQLVTLN
 651 VMNQQLPLTG TSLVLSQSSY NLSF*

The cp7107 nucleotide sequence <SEQ ID 126> is:

1 ATGAGTATAG TCAGAAATTC TGCATTGCCA CTTCCTGTGT TAAGCAGATC
 51 CGAAACCTTT AAAAAAGTTA GGTGCGCATAT GAAATTTATG AAAGTCTTTA
 101 CTCCATGGAT TTATCGAAAA GATCTTTGGG TAACAGCATT CTTACTGACA
 151 GCAATTCAG GATCTTTTGC ACATACTCTT GTTGATATAG CAGGAGAACC
 201 TCGGCATGCT GCTCAAGCAA CAGGAGTTTC TGGAGATGGT AAAATTGTTA
 251 TAGGAATGAA AGTTCCGGAT GATCCTTTTG CTATAACTGT AGGATTTCAA
 301 TATATTGATG GGCATTTGCA ACCCTTAGAG GCAGTACGTC CTCAATGCTC
 351 TGTATACCCT AATGGTATAA CCCCAGACGG AACGGTTATT GTGGGTACAA
 401 ACTATGCCAT CGGGATGGGT AGTGTGTGCTG TGAAATGGGT AAATGGCAAG
 451 GTTTCTGAAC TTCCCAGCT CCCTGACACC CTCGATTCTG TAGCATCGGC
 501 AGTTTCTGCA GATGGAAGAG TGATTGGAGG GAATAGAAAT ATAAATCTTG
 551 GCGCTTCTGT TGCTGTGAAA TGGGAGGACG ACGTGATTAC ACAACTCTCT
 601 TCTCTTCTCTG ATGCTATGAA TGCTTGTGTT AACGGAATTT CTTCAGATGG

5 651 TTCTATAATT GTAGGAACCA TGGTAGACGT GTCATGGAGA AATACCGCAG
 701 TACAATGGAT CGGGGATCAG CTCTCTGTTA TTGGGACTTT AGGAGGAACT
 751 ACTTCTGTTG CTAGTGCAAT CTCAACAGAT GGCAGTGTGA TTGTAGGAGG
 801 TTCTGAAAT GCAGATTCTC AGACTCATGC CTATGCTTAT AAAAACGGTG
 851 TTATGAGCGA TATAGGGACC CTCGGAGGTT TTTATTCTTT AGCACATGCA
 901 GTATCTTCAG ATGGTTCTGT GATTGTAGGA GTATCCACGA ACTCTGAGCA
 951 TAGATATCAT GCATTCCAAT ATGCTGATGG ACAGATGGTA GATTTAGGAA
 1001 CTTTAGGAGG GCCTGAATCT TATGCTCAAG GTGTGTCTGG AGATGGAAAG
 1051 GTAATTGTGG GTAGAGCACA AGTACCATCT GGAGATTGGC ATGCGTTCCT
 1101 ATGTCCTTTC CAAGCTCCGA GCCCTGCTCC TGTCCATGGG GGAAGCACTG
 1151 TCGTAAC TAG CCAGAATCCA CGTGGAAATGG TAGATATCAA TGCTACGTAC
 1201 TCCTCTTTGA AAAATAGCCA ACAACAATA CAAAGATTGC TTATCCAGCA
 1251 TAGTGCAAAA GTTGAAGTG TATCCTCAGG AGCACCATCT TTTACAAGTG
 1301 TGAAAGGTGC GATCTCAAAA CAGAGCCCTG CAGTGCAAAA TGATGTACAG
 1351 AAAGGGACGT TTTTAAAGTTA CCGTTCCTCA GTTCATGGAA ACGTGCAGAA
 1401 TCAGCAATTG CTCACAGGAG CTTTTATGGA CTGGAACTC GCTTCAGCTC
 1451 CTAAATGCGG CTTTAAAGTA GCTCTCCACT ATGGCTCTCA AGATGCTCTC
 1501 GTAGAACGTG CAGCTCTTCC TTACACAGAA CAAGGCTTAG GAAGCAGTGT
 1551 CTTGTCAGGT TTTGGAGGAC AAGTTCAAGG ACGCTATGAC TTTAATTTAG
 1601 GAGAAACTGT TGTCTGCAA CCCTTATGG GCATTCAAGT TCTCCACCTA
 1651 AGTAGAGAAG GGTATTCTGA GAAGAATGTT CGATTTCCTG TAAGCTATGA
 1701 TTCTGTAGCC TACTCAGCAG CTACTAGCTT TATGGGTGCG CATGTATTTG
 1751 CCTCCCTAAG CCCTAAAATG AGTACAGCAG CAACTTTAGG TGTGGAGAGA
 1801 GATCTGAATT CACATATAGA TGAATTTAAG GGATCCGTCT CTGCTATGGG
 1851 AACTTTGTC TTGGAAAATT CTACAGTGAG TGTTTTAAAGA CCTTTTGCTT
 1901 CTCTTGCTAT GTACTATGAC GTAAGACAAC AGCAACTCGT GACGTGTGCA
 1951 GTAGTTATGA ATCAACAACC CTTAACAGGC ACACTAAGCT TAGTAAGCCA
 2001 AAGTAGCTAT AATCTTAGCT TCTAA

The PSORT algorithm predicts an inner membrane location (0.100).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 63A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 63B) and FACS (Figure 63C) analyses.

These experiments show that cp7107 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 64

The following *C.pneumoniae* protein (PID 4376467) was expressed <SEQ ID 127; cp6467>:

1 **MLRFFAVFLS** **TLWLITSGCS** PSQSSKGIFV VNMKEMPRSL DPGKTRLIAD
 51 QTLMRHLYEG LVEEHSQNGE IKPALAESYT ISEDGTRYTF KIKNILWSNG
 101 DPLTAQDFVS SWKEILKEDA SSVLYAFLEP IKNARAIFDD TESPENLGVR
 140 151 ALDKRHLEIQ LETPCAHLFH FLTLPFFPV HETLRNYSTS FEEMPITCGA
 201 FRPVSLEKGL RLHLEKNPMY HNSRVKLHK IIVQFISNAN TAILFKHKK
 251 LDWQGPWGE PIPPEISASL HQDDQLFSLP GASTTWLLFN IQKKPWNNAK
 301 LRKALSLAID KDMLTKVVYQ GLAEPTDHIL HPRLYPGTYP ERKRQNERIL
 351 EAQQLFEEAL DELQMTREDL EKETLTFSTF SFSYGRICQM LREQWKKVLK
 401 FTIPIVGQEF FTIQKNFLEG NYSLTVNQWT AAFIDPMSYL MIFANPGGIS
 45 451 PYHLQDSHFQ TLLIKITQEH KKHRLRNQLII EALDYLEHCH ILEPLCHPNL
 501 RIALNKNIKN FNLFVRRTSD FRFIEKL*

A predicted signal peptide is highlighted.

The cp6467 nucleotide sequence <SEQ ID 128> is:

50 1 ATGCTCCGTT TCTTCGCTGT ATTTATATCA ACTCTTTGGC TCATTACCTC
 51 AGGATGTGCC CCATCCCAAT CCTCTAAAGG AATTTTGTG GTAAATATGA
 101 AGGAAATGCC ACGCTCCTTG GATCCTGGAA AACTCGTCT CATTCGAGAC
 151 CAAACTCTAA TCGGTCATCT ATATGAAGGA CTCGTCCAG AACATTCCTCA
 201 AAATGGAGAG ATTAACCAG CCCTTGCAGA AAGCTACACC ATCTCCGAG
 55 251 ACGGGACTCG GTACACATTT AAAATCAAAA ACATCCTTTG GAGTAACGGA
 301 GACCCCTCTGA CAGCTCAAGA CTTTGTCTCC TCTTGAAGG AAATCCTAAA

5
10
15
20
25

```

351 GGAAGATGCG TCCTCCGTAT ATCTCTATGC GTTTTACCT ATCAAAAATG
401 CTCGGGCAAT CTTTGATGAT ACTGAGTCTC CAGAAAATCT AGGAGTCCGA
451 GCTTTAGATA AGCGTCATCT CGAAATTCAG TTAGAAACTC CTGCGCGCA
501 TTTCTACAT TTCTGACTC TTCTATTTT TTTCCCTGTT CATGAAACTC
551 TGCGAAACTA TAGCACCTCT TTTGAAGAGA TGCCCATTAC CTGCGGTGCT
601 TTCCGCCCTG TGTCTCTAGA AAAAGGCCTG AGACTCCATC TAGAGAAAAA
651 CCCTATGTAC CATAATAAAA GCCGTGTGAA ACTACATAAA ATTATTGTAC
701 AGTTTATCTC AAACGCTAAC ACTGCAGCCA TTCTATTCAA ACATAAGAAA
751 TTAGATTGGC AAGGACCTCC TTGGGGAGAA CCTATCCCTC CAGAAATCTC
801 AGCTTCTCTA CATCAAGATG ACCAGCTCTT TTCCTTCCG GCGCTTCGA
851 CTACATGGTT ACTCTTTAAT ATACAAAAAA AACCTTGGAA CAATGCTAAA
901 TTACGCAAGG CATTGAGCCT TGCAATAGAC AAAGATATGT TAACCAAAGT
951 GGTATACCAA GGTCTTGACG AACCTACAGA TCATATCCTA CATCCAAGAC
1001 TTTATCCAGG GACCTATCCC GAACGGAAAA GACAAAACGA AAGAATTCTT
1051 GAGGCTCAAC AACTCTTTGA AGAAGCTCTA GACGAACTTC AAATGACACG
1101 CGAAGATCTA GAAAAGGAAA CTTTGACTTT CTCACCTTT TCTTTTCTT
1151 ACGGAAGGAT TTGCCAAATG CTAAGAGAAC AATGGAAGAA AGTCTTAAAA
1201 TTTACTATCC CTATAGTAGG CCAAGAGTTT TTCACAATAC AAAAAAACTT
1251 CCTAGAGGGG AACTATTCCC TAACCGTGAA CCAATGGACC GCAGCATTTA
1301 TTGATCCGAT GTCTTATCTC ATGATCTTTG CCAATCCTGG AGGAATTTC
1351 CCCTATCACC TCCAAGATTC ACACTTTCAA ACTCTTCTCA TAAAGATCAC
1401 TCAAGAACAT AAAAAACACC TACGAAATCA GCTTATTATT GAAGCCCTTG
1451 ACTATTTAGA ACACCTTCAC ATTCTCGAAC CACTATGTCA TCCAAATCTT
1501 CGAATTGCTT TGAACAAAAA CATTAATAAC TTAAATCTTT TTGTTGACG
1551 AACTTCAGAC TTTCGTTTAA TAGAAAAACT ATAG

```

The PSORT algorithm predicts an outer membrane lipoprotein (0.790).

30 The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion protein, as shown in Figure 64A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 64B). The recombinant GST-fusion protein was also used to immunise mice, whose sera were used in a Western blot (Figure 64C) and for FACS analysis (Figure 64D).

These experiments show that cp6467 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 65

35 The following *C.pneumoniae* protein (PID 4376679) was expressed <SEQ ID 129; cp6679>:

40

```

1 MRKMLVLLAS LGLLSPTLSS CTHLGSSGSY HPKLYTSGSK TRGVIAMLPV
51 FHRPGKSLEP LPWNLQGEFT EEISKRFYAS EKVFLIKHNA SPQTVSQFYA
101 PIANRLPETI IEQFLPAEFI VATELLEQKT GKEAGVDSVT ASVRVRVFDI
151 RHHKIALIQ EIIECSQPLT TLVNDYHRYG WNSKHFDSTP MGLMHSRLFR
201 EVVARVBGYV CANYS*

```

A predicted signal peptide is highlighted.

The cp6679 nucleotide sequence <SEQ ID 130> is:

45
50

```

1 ATGCGAAAAA TGTTGGTATT ATTGGCATCT TTAGGACTTC TATCCCCAAC
51 CCTATCCAGC TGCACCTACT TAGGCTCTTC AGGAAGTTAT CATCCTAAGC
101 TATACACTTC AGGGAGCAAA ACTAAAGGTG TGATTGCGAT GCTTCTGTGA
151 TTTTCATCGCC CAGGAAAGAG TCTTGAACCT TTACCTTGGG ACCTCCAAGG
201 AGAATTTACT GAAGAGATCA GCAAAGGTTT TTATGCTTCG GAAAAGGTCT
251 TCCTGATCAA GCACAATGCT TCACCTCAGA CAGTCTCTCA GTTCTATGCT
301 CCGATTGCGA ATCGTCTACC CGAAACAATT ATTGAGCAAT TTCTTCTGTC
351 AGAATTCATT GTTGTACAG AACTGTTAGA ACAAAGACA GGGAAAGAAG
401 CAGGTGTCGA TTCTGTAACA GCGTCTGTAC GTGTTGCGCT TTTTGATATC
451 CGTCATCATA AAATAGCTCT CATTTATCAA GAGATTATCG AATGCAGCCA
501 GCCTTTAACT ACCCTAGTCA ATGATATCA TCGCTATGGC TGGAACTCAA
551 AACATTTTGA TTCAACGCC ATGGGCTTAA TGCATAGCCG TCTTTTCCCG

```

601 GAAGTTGTTG CCAGAGTTGA GGGCTATGTT TGTGCTAACT ACTCGTAG

The PSORT algorithm predicts an inner membrane location (0.149).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 65A) and as a GST-fusion product (Figure 65B). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 65C) and for FACS analysis.

These experiments show that cp6679 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 66

The following *C.pneumoniae* protein (PID 4376890) was expressed <SEQ ID 131; cp6890>:

```

10      1  MKQLLFCVCV FAMSCSAYAS PRRQDPSVMK ETFRNNYGII VSGQEWVKRG
      51  SDGTITKVLK NGATLHEVYS GLLHGEITL TFPHTTALDV VQIYDQGRLV
     101  SRKTFVNGL PSQEELFNEG GTFVLTRWPD NNDSDTITKP YFIETTYQGH
     151  VIEGYSYTSFN GKYSSSIHNG EGVRSVFSNN NILSEETFN EGVMVKYTF
     201  YPNRDPESIT HYONGQPHGL RLTYLQGGIP NTIEEWRYGF QDGTITVFKN
     251  GCKTSEIAYV KGVKEGLELR YNEQEIVAEV VSWRNDFLHG ERKIYAGGIQ
     301  KHEWYYRGRS VSKAKFERLN AAG*
  
```

A predicted signal peptide is highlighted.

The cp6890 nucleotide sequence <SEQ ID 132> is:

```

20      1  ATGAAACAAT TACTTTTCTG TGTTCGCGTA TTTGCTATGT CATGTTCTGC
      51  TTACGCATCC CCACGACGAC AAGATCCTTC TGTATGAAG GAAACATTCC
     101  GAAATAATTA TGGCATTATT GTTCCCGGTC AAGAATGGGT AAAGCGTGGT
     151  TCTGACGGCA CCATCACCAA AGTACTCAA AATGGAGCTA CCCTGCATGA
     201  AGTTTATCTT GGAGGCCTCC TTCATGGGGA AATTACCTTA ACGTTTCCCC
     251  ATACCACAGC ATTGGACGTT GTTCAAATCT ATGATCAAGG TAGACTCGTT
     301  TCTCGCAAAA CCTTTTGTGT GAACGGTCTT CCATCTCAAG AAGAGCTGTT
     351  CAATGAAGAT GGCACGTTTG TCCTCACACG ATGGCCGGAC AACACGACA
     401  GTGATACCAT CACAAAGCCT TACTTCATAG AAACGACATA TCAAGGGCAT
     451  GTCATAGAAG GAAGTTATAC TTCCTTTAAT GGGAAATACT CCTCATCCAT
     501  CCACAATGGA GAGGGAGTTC GTTCTGTGTT CTCCTCCAAT AACATCCTTC
     551  TTTCTGAAGA GACCTTCAAT GAAGGTGTCA TGGTGAAATA TACCACATTC
     601  TATCCGAATC GCGATCCCGA ATCGATTACT CATATCAAA ATGGACAGCC
     651  TCACGGCTTA CGGCTAACAT ATCTACAAGG TGGCATCCCC AATACGATAG
     701  AGGAGTGGCG TTATGGCTTT CAAGACGGAA CGACCATCGT ATTTAAAAAT
     751  GGTGTGAAGA CATCTGAGAT CGCTTATGTT AAGGGAGTGA AAGAAGGTTT
     801  AGAACTCGCG TACAATGAAC AGGAAATTGT AGCTGAAGAA GTTCTTTGGC
     851  GTAATGATTT TCTGCATGGA GAACGTAAGA TCTATGCTGG AGGAATCCAA
     901  AAGCATGAAT GGTATTACCG CGGGAGATCT GTATCTAAAG CCAAATCCGA
     951  GCGGCTAAAT GCTGCAGGAT AG
  
```

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 66A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 66B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6890 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 67

The following *C.pneumoniae* protein (PID 6172323) was expressed <SEQ ID 133; cp0018>:

1 MKTSVSMLLA LLCSGASSIV LHAATTPLPNP EDGFIEGNT NTFSPKSTTD
 51 AAGTTYSLTG EVLYIDPGKG GSITGTCFVE TAGDLTFLGN GNTLKFSLVD
 101 AGANIAVAHV QGSKNLSFTD FLVLVITESP KSAVTTGKGS LVSLGAVQLQ
 151 DINTLVLTSN ASVEDGGVIK GNSCLIQGIK NSAIFGQNTS SKKGGAISTT
 201 QGLTIENNLG TLKFENENKAV TSGGALDLGA ASTFTANHEL IFSQNKTSNG
 251 AANGGAINCS GDLTFTDNTS LLLQENSTMQ DGGALCSTGT ISITGSDSIN
 301 VIGNTSGQKG GAISAASLKI LGGQGGALFS NNVTTHATPL GGAIFINTGG
 351 SLQLFTQGGD IVFEGNQVTT TAPNATTKRN VIHLESTAKW TGLAASQGNA
 401 IYFYDPITFN DTGASDNLRI NEVSANQKLS GSIVFSGERL STAEIAENL
 10 451 TSRINQPVTL VEGSLVLKQG VTLITQGFSSQ EPESTLLLDL GTSL*

A predicted signal peptide is highlighted.

The cp0018 nucleotide sequence <SEQ ID 134> is:

1 ATGAAGACTT CAGTTTCTAT GTTGTGGGCC CTGCTTTGCT CGGGGGCTAG
 51 CTCTATTGTA CTCCATGCCG CAACCACTCC ACTAAATCCT GAAGATGGGT
 15 101 TTATTGGGGA GGGCAATACA AATACTTTT CTCCGAAATC TACAACGGAT
 151 GCTGCAGGAA CTACCTACTC TCTCACAGGA GAGGTTCTGT ATATAGATCC
 201 GGGGAAAGGT GGTTCATTA CAGGAACCTG CTTTGTAGAA ACTGCTGGCG
 251 ATCTTACATT TTTAGGTAAT GGAAATACCC TAAAGTCTCT GTCGCTAGAT
 301 GCAGGTGCTA ATATCGCGGT TGCTCATGTA CAAGGAAGTA AGAATTTAAG
 351 CTTCACAGAT TCTCTTCTC TGGTGATCAC AGAATCTCCA AAATCGCTG
 401 TTAATACAGG AAAAGGTAGC CTAGTCAGTT TAGGTGCAGT CCAACTGCAA
 451 GATATAAACA CTCTAGTTCT TACAAGCAAT GCCTCTGTCT AAGATGGTGG
 501 CGTGATTAAA GGAACTCCT GCTTGATTCA GGAATCAAA AATAGTGCAG
 551 TTTTGGGACA AAATACATCT TCGAAAAAAG GAGGGGCGAT CTCCACGACT
 25 601 CAAGGACTTA CCATAGAGAA TAACTTAGGG ACGCTAAAGT TCAATGAAAA
 651 CAAAGCAGTG ACCTCAGGAG GCGCCTTAGA TTTAGGAGCC GCGTCTACAT
 701 TCACTGCGAA CCATGAGTTG ATATTTTCAC AAAATAAGAC TTCTGGGAAT
 751 GCTGCAAATG GCGGAGCCAT AAATTGCTCA GGGGACCTTA CATTTACTGA
 801 TAACACTTCT TTGTTACTTC AAGAAAATAG CACAATGCAG GATGGTGGAG
 30 851 CTTTGTGTAG CACAGGAACC ATAAGCATT CCGGTAGTGA TTCTATCAAT
 901 GTGATAGGAA ATACTTCAGG ACAAAGGGA GGAGCGATTT CTGCAGCTTC
 951 TCTCAAGATT TTGGGAGGGC AGGGAGGCGC TCTCTTTCT AATAACGTAG
 1001 TGACTCATGC CACCCCTCTA GGAGGTGCCA TTTTATCAA CACAGGAGGA
 1051 TCCTTGACAG TCTTCACTCA AGGAGGGGAT ATCGTATTCG AGGGGAATCA
 35 1101 GGTCACTACA ACAGCTCCAA ATGCTACCAC TAAGAGAAAT GTAATTCACC
 1151 TCGAGAGCAC CGCGAAGTGG ACGGGACTTG CTGCAAGTCA AGGTAACGCT
 1201 ATCTATTCTT ATGATCCCAT TACCACCAAC GATACGGGAG CAAGCGATAA
 1251 CTTACGTATC AATGAGGTCA GTGCAATCA AAAGCTCTCG GGATCTATAG
 1301 TATTTTCTGG AGAGAGATTG TCGACAGCAG AAGCTATAGC TGAAAATCTT
 40 1351 ACTTCGAGGA TCAACCAAGC TGTCACCTTA GTAGAGGGGA GCTTAGTACT
 1401 TAAACAGGGA GTGACCTGA TCACACAAGG ATTCTCGCAG GAGCCAGAAT
 1451 CCACGCTTCT TTTGGATCTG GGGACCTCAT TATAA

The PSORT algorithm predicts outer membrane (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 67A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 67B) and for FACS analysis.

These experiments show that cp0018 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 68

50 The following *C.pneumoniae* protein (PID 4376262) was expressed <SEQ ID 135; cp6262>:

1 MRKLRLIAIV LIALSIILIA GGVLLTVAI PGLSSVISSP AGMGACALGC
 51 VMLALGIDVL LKKREVPVL ASVTTTPGTG SPRSGISISG ADSTIRSLPT
 101 YLLDEGHPQS MRKLRLIAIV LIVFSIILIA SGVLLTVAI PGLSSVISSP
 151 AGMGACALGC VMLALGIDVL LKKREVPVL ASVTTTPGTG SPRSGISISG
 201 ADSTIRSLPT YPLDEGHPQS MRKLRLIAIV LIVFSIILIA SGVLLTVAI
 251 PGLSSIISP AEMGACALGC VMLALGIDVL LKKREVPVL PAPIPEEVVI

5
10
15

```

301 DDIDEESIRL QQEBAEALAR LPEEMSAFEG YIKVVESHLE NMKSLPYDGH
351 GLEEKTKHQI RVVRSSLKAM VPEFLDIRRI FEEEEFFFLS ARKRLIDLAT
401 TLVERKILTE QLERNNLRKA FSYLYQDSIF KKIIDNFEKL AWKFMILSKS
451 ICRFTIIFEN HEHGVAKSL L HKNVLLLEKV TYRSLQKSYR DIGMSSAKMK
501 ILHGNPFFSL EDNKKTIMKE HAEMLESLS YRKVFLALSD ENVVDTPSDP
551 KKWDLSGIPC RDALSEISRD EQWQKKAHLK HQESLYTQAR DRLTDQSSKE
601 NQKELEKAEQ EYISSWERYK KFEIERVQER IRAIQKLYPN ILEREETTGT
651 QETVTPTVQG TTASSDLTDI LGRIEVSSRE DNQONQESCVK VLRSEHEVMS
701 WEVKQYGPKE KKEFDQMGSL LERFFTEHIE ELEVLOKDYS KHLSTYFKKVN
751 NKKEVQYAKF RLKVLSDLE GILAQTESAE SLLTQBELPI LATRGALEKA
801 VFKGSLCCAL ASKAKPYFEE DPRFQSDSTQ LRALTIRLQE AKASLEEEIK
851 RFSNLENDIA EERLLKESK QTFERAGLV LREIAVESTY DLRLSTNTWE
901 GTPSEKVVYF SMYLNYNEE KRRAKTRLVE MTQRYRDFKM ALEAMQFNEE
951 ALLQEELSIQ APSE*

```

15 A predicted signal peptide is highlighted.

The cp6262 nucleotide sequence <SEQ ID 136> is:

20
25
30
35
40
45
50
55
60
65

```

1 ATGAGGAAC TTCGTATTCT TCGCATCGTT CTCATAGCTT TGAGCATTAT
51 TTTGATTGCA GGTGGTGTGG TATTGCTTAC TGTAGCGATC CCTGGATTAA
101 GTTCAGTCAT TTTTCCCCG GCAGGGATGG GTGCCTGTGC TTTGGGATGT
151 GTGATGCTTG CTTTAGGGAT CGATGTTCTT CTGAAGAAAC GAGAAGTCCC
201 TATAGTTCTC GCATCTGTAA CTACGACACC AGGAACTGGC AGCCCTAGAA
251 GTGGTATTTC TATTTCAGGA GCTGATAGCA CCATACGTTT TCTTCTTACG
301 TATCTCTTGG ACGAGGGACA TCCACAATCC ATGAGGAAC TTCGTATTCT
351 TCGCATCGTT CTCATAGTTT TTAGCATTAT TTTGATTGCA AGTGGTGTGG
401 TATTGCTTAC TGTAGCGATC CCTGGATTAA GTTCAGTCAT TTTTCCCCG
451 GCAGGGATGG GTGCCTGTGC TTTGGGATGT GTGATGCTTG CTTTAGGGAT
501 CGATGTTCTT CTGAAGAAAC GAGAAGTCCC TATAGTTCTC GCATCTGTAA
551 CTACGACACC AGGAACTGGC AGCCCTAGAA GTGGTATTTC TATTTCAGGA
601 GCTGATAGCA CCATACGTTT TCTTCTTACG TATCCCTTGG ACGAGGGACA
301 TCCACAATCC ATGAGGAAC TTCGTATTCT TCGCATCGTT CTCATAGTTT
701 TTAGCATTAT TTTGATTGCA AGTGGTGTGG TATTGCTTAC TGTAGCGATC
751 CCTGGATTAA GCTCGATCAT TTCTTCCCCA GCGGAGATGG GTGCTTGTGC
801 TTTGGGATGT GTGATGCTTG CTTTGGGGAT CGACGTTCTT CTGAAGAAAC
851 GAGAAGTCCC TATAGTAGTT CCCGCACCTA TTCTGAAGA AGTCGTCATA
351 GATGATATAG ATGAAGAGAG TATACGGCTG CAGCAGGAAG CTGAAGCCGC
901 TTAGCAAGA CTTCTTGAGG AGATGAGTGC ATTTGAAGGT TACATAAAG
951 TTGTCGAGAG TCATTTGGAG AACATGAAA GCCTGCCCTA TGATGGTCAT
1001 GGGCTAGAAG AGAAAACGAA ACATCAGATA AGAGTCGTCA GATCTTCTTT
1051 GAAGGCTATG GTTCCAGAA TTTTAGATAT CAGAAGAATT TTGAAGAAG
1101 AAGAGTTCTT TTTTCTCTCA GCTCGCAAAC GACTTATAGA TTAGCTACT
1151 ACTTTAGTAG AGAGAAAAT TTTAACAGAG CAACCTGAGC GCAATAAATT
1201 AAGGAAAGCG TTTTCTTATT TATATCAGGA CTCAATTTT AAAAAAATTA
1251 TTGATAACTT CGAGAAGTTA GCATGGAAAT TTATGATTTT GAGTAAATCA
1301 ATTTGTCGAT TTACAATTAT TTTTGAAAAT CATGAACATG GTGTAGCAAA
451 GAGCCTGTTA CACAAGAAAT CAGTGTACT GGAGAAGGTA ATCTATAGGA
1401 GTTTGCAAAA AAGCTATAGA GATATAGGCA TGTCTCTGTC AAAGATGAAA
1451 ATCTTGCACG GCAACCCTTT TTTCTCTTTG GAAGATAATA AAAAGACGAT
1501 AATGAAAGAA CACGCAGAGA TGCTTGAAAG TCTCAGTAGC TATAGGAAGG
1551 TATTTTTAGC TCTATCTGAT GAGAACGTTG TAGATACACC TAGCGATCCA
1601 AAGAAATGGG ATTTGTCAGG AATCCCTGT AGGGACGCGT TGTCTGAGAT
1651 TTTCTCGTAT GAACAGTGGC AGAAGAAAGC ACATCTAAAG CATCAAGAGT
1701 CCTCTATAC GCAAGCTAGG GATCGTTTAA CAGACCAGAG CTCTAAAGAA
1751 AATCAGAAAG AGTTAGAGAA AGCTGAACAA GAGTACATAT CTTCTTGGGA
1801 ACGGGTTAAA AAATTTGAGA TTGAGAGAGT ACAGGAGAGG ATACGGGCAA
1851 TTCAAAGCT TTATCCTAAT ATCCTCGAGA GAGAAGAAGA AACCACAGGT
55 1901 CAGGAGACTG TGACTCCAAC TGTTCAAGGG ACGACGGCTT CATCCGATTT
1951 AACAGATATT TTAGGAAGAA TAGAGGTCTC CAGTAGGGAG GATAATCAGA
2001 ATCAAGAGTC TTGTGTAAAA GTCTTAAGAA GTCATGAGGT AGAAATGAGC
2051 TGGGAAGTCA AACAGAGTA TGGCCCTAAG AAAAAAGAA TTCAGGATCA
2101 AATGGGTCTT TTAGAGAGGT TTTTACAGA GCATATTGAA GAGTTAGAAG
2151 TATTACAGAA GGACTACTCT AAACACTTGT CTTATTTTAA AAAAGTAAAC
2201 AATAAGAAAG AGGTTCATA TGCGAAGTTT AGGTTGAAGG TTTTAGAGTC
2251 AGATTTAGAA GGGATTCTAG CTCAGACTGA GAGTGCTGAG AGTCTGTTAA
2301 CTCAAGAAAG ACTTCCGATT CTTGCAACTC GGGGAGCCTT AGAGAAAGCT
2351 GTTTTCAAAG GGAGTCTATG TTGCGCGCTA GCAAGCAAAG CAAAACCCTA

```

5
2451 TTTTGAAGAG GATCCCAGAT TCCAAGATTC TGATACGCAA TTGCGAGCTC
2501 TGA CTCTAAG GTTACAGGAG GCTAAGGCAA GCCTGGAAGA AGAGATAAAG
2551 AGATTTTCAA ATCTTGAGAA CGATATTGCA GAGGAAAGAC GCCTTCTTAA
2601 AGAGAGCAAG CAGACGTTTCG AAAGAGCAGG TTTAGGGGTT CTCCGAGAAA
2651 TTGCAGTCGA GTCTACTTAT GATTTCGCTT CCTTAACAAA TACATGGGAA
2701 GGGACCCAG AGAGTGAGAA GGTCTATTTT AGCATGTATC TTAATTATTA
2751 CAACGAAGAG AAACGTAGGG CTAAAACAAG ATTGGTTGAA ATGACACAGA
2801 GGTATAGAGA TTTTAAAATG GCCTTGGAAG CTATGCAGTT TAATGAAGAA
2851 GCCCTTTTGC AAGAGGAAC TCTATTCAA GCTCCAGTG AATAA

10 The PSORT algorithm predicts inner membrane (0.660).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 68A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 68B) and for FACS analysis.

15 These experiments show that cp6262 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 69

The following *C.pneumoniae* protein (PID 4376269) was expressed <SEQ ID 137; cp6269>:

20
1 MYQENLRLLLE RLLYSNVQKS YADRLFSYEK TKMVHDTPLI PWEEDKEKCA
51 EAEKAFLEQQ KILLDYGKSI FWLNENDEIN LNDPWSWGLN TVRTRKVFQE
101 VDDSERWNHK VLIQKLEDDY EKLLEESSKE STEANKKLLS DLVDRLEDAK
151 TKFFLKKQEE VETRVKDLRA RYGGTVDPKQ DTEAKKKVEL EASLETFLDS
201 IESELVQCLE DQDIYWKEQD VKDLARTQEL EEQDIEAKRE EAAEDLRLSLN
251 ERLKKSXTML DRKWHIENA EDSITWWTSS IEMKDMKARL KILKEDITSV
301 LPEIDEIETC LSLEELPLL TRELITKSYL KFKICSETLL KMTSVFENNI
25 YVQEYEVQLQ NLGFKLQGIS QRFGKKQDDF ANLEEQVALQ KKRLRELTON
401 FEIQGFNFMK EDFKAAAKDL YIRSTAEQKM NFDVPCMELE RRYHEEVNKP
451 LLELMYNCAD SYRDAKKKLC SLRLDEKELL QKEIKKEEFY QKKQQRHADR
501 SRHTTYQKLR IAEELALELK KKI*

The cp6269 nucleotide sequence <SEQ ID 138> is:

30
1 ATGTACCAGG AGAATCTAAG ATTGTTGGAA AGGCTTCTTT ATAATAGTGT
51 TCAAAGAGGC TATGCGGATC GGCTGTTTTC CTATGAAAAG ACAAAGATGG
101 TGCACGATAC TCCGCTGATT CCTTGGGAAG AGGATAAGGA AAAATGTGCT
151 GAAGCTGAGA AAGCTTTCTT AGAGCAACAG AAGATTCTCC TAGATTATGG
35 AAAATCTATC TTTTGGCTGA ATGAGAACGA TGAGATCAAT TTAAACGATC
251 CTGAGAGTTG GGGTCTTAAT ACGGTGAGGA CTAGGAAAGT ATTCCAAGAG
301 GTTGACGACA GTGAACGTTG GAATCATAAG GTACTCATT CAAAACCTCGA
351 GGACGATTAT GAGAAACTTC TAGAGGAAAG TTCAAAGAG TCTACTGAAG
401 CAAATAAGAA GCTTTTATCT GACTTAGTAG ATCGTCTTGA AGATGCTAAG
451 ACAAATTTT TCCTGAAGAA ACAGGAGGAG GTGGAGACTC GCGTTAAGGA
501 TCTTAGAGCT CGATATGGAG GCACAGTAGA TCCTAAGCAG GATACGGAAG
551 CTAAGAAGAA AGTCGAATTG GAGGCTAGCT TAGAAACCTT TTTAGATTCC
601 ATCGAATCAG AGCTAGTACA GTGTTTAGAA GATCAAGATA TATATTGGAA
651 AGAACAGGAT GTCAAAGATC TAGCACGTAC GCAAGAGCTC GAGGAACAAG
701 ATATTGAAGC GAAGAGGGAA GAAGCTGCCG AAGACCTAAG AAGTCTTAAT
45 GAGCGTTTAA AGAAGTCAAA AACTATGTTA GATAGGGCTA AATGGCATAT
751 TGAAAATGCT GAGGACAGTA TTACCTGGTG GACTAGTCAG ATAGAAATGA
801 AGGATATGAA AGCAAGACTG AAGATCTTAA AAGAAGATAT AACAAAGTGT
851 CTACCTGAAA TAGATGAGAT TGAACGTTG TTAAGCTTAG AGGAGCTTCC
901 TTTGCTTACG ACCAGGGAAC TCTTAACTAA GTCCTACCTA AAGTTTAAGA
50 1001 TTTGTTTCGGA AACACTATTA AAAATGACTT CTGTGTTTGA GAACAATATC
1051 TATGTTTACG AGTACGAGGT TCAGCTGCAA AATCTAGGGT TTAAGTTACA
1101 AGGTATATCT CAGAGATTCTG GAAAGAAACA AGACGATTTT GCGAATCTAG
1151 AGGAACAGGT TGCTTTGCAA AAGAAACGAC TCAGAGAGCT CACTCAGAAT
1201 TTTGAAATAC AAGGATTCAA TTTTCATGAAA GAAGATTTTA AGGCAGCCGC
55 1251 TAAAGATCTT TATATAAGAA GTACAGCTGA ACAAAGATG AACTTTGATG
1301 TGCCTTGCAT GGAGCTCTTC CGTAGGTATC ATGAGGAGGT CAACAAGCCG
1351 CTTCTTGAGT TGATGTACAA TTGTGCAGAC AGTTATAGAG ATGCTAAGAA

1401 AAAGCTTTGC TCTCTACGTC TTGATGAAAA AGAGTTATTA CAAAAAGAAA
 1451 TCAAGAAAGA GGAATTTTAT CAAAAGAAAC AACAAAGGCA TGCAGATAGA
 1501 TCACGTGATA CTACGTATCA AAAGCTACGA ATTGCTGAAG AGCTTGCTCT
 1551 TGAGCTGAAG AAGAAAATCT AA

- 5 The PSORT algorithm predicts cytoplasmic location (0.412).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 69A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 69B) and for FACS analysis.

- 10 These experiments show that cp6269 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 70

The following *C.pneumoniae* protein (PID 4376270) was expressed <SEQ ID 139; cp6270>:

1 MKIPLRLLI SLVPTLSMSN LLGAATTEEL SASNSFDGTT STTSFSSKTS
 51 SATDGTNYVF KDSVVIENVP KTGETQSTSC FKNDAAGDL NFLGGGFSFT
 101 FSNIDATTAS GAAIGSEAN KVTTLSGFSA LSFLKSPAST VTNGLGAINV
 151 KGNLSLLDND KVLIQDNFST GDGGAINCAG SLKIANNKSL SFIGNSSSTR
 201 GGAIHTKNLT LSSGGETLFQ GNTAPTAAGK GGAIAIADSG TLSISGDSGD
 251 IIFEGNTIGA TGTVSHSAID LGTSAKITAL RAAQGHITIYF YDPITVTGST
 301 SVADALNINS PDTGDNKEYT GTIVFSGEKL TEAEAKDEKN RTSKLLQNV
 351 FKNGTVVLKG DVVLSANGFS QDANSKLIMD LGTSLVANTE SIELTNLEIN
 401 IDSLRNGKKI KLSAATAQKD IRIDRPVULA ISDESFYQNG FLNEDHSYDG
 451 ILELDAGKDI VISADSRSID AVQSPYGYQG KWTINWSTDD KKATVSWAKQ
 501 SFNPPTAEQEA PLVFNLLWGS FIDVRSFQNF IELGTEGAPY EKRFVWAGIS
 551 NVLHRSGREN QRKFRHVSOG AVVGASTRMP GGDTLGLGFA QLFARDKDYF
 601 MNTNFAKTYA GSLRLQHDAS LYSVVSILLG EGGLREILLP YVSKTLPSCF
 651 YGQLSYGHTD HRMKTESLPP PPPTLSTDHT SWGGYVWAGE LGTRVAVENT
 701 SGRGFFQEYT PFVKVQAVYA QDSFVELGA ISRDFSDSHL YNLAIPLGIK
 751 LEKRFABQYY HVVAMYSQDV CRSNPKCTTT LLSNQGSWKT KGSNLARQAG
 801 IVQASGFRSL GAAELFGNF GFIEWRGSRS YNVDAGSKIK F*

- 30 A predicted signal peptide is highlighted.

The cp6270 nucleotide sequence <SEQ ID 140> is:

1 ATGAAGATTC CACTCCGCTT TTTATGATA TCATTAGTAC CTACGCTTTT
 51 TATGTCGAAT TTATTAGGAG CTGCTACTAC CGAAGAGTTA TCGGCTAGCA
 101 ATAGCTTCGA TGGAACTACA TCAACAACAA GCTTTTCTAG TAAACATCA
 151 TCGGCTACAG ATGGCACCAA TTATGTTTTT AAAGATTCTG TAGTTATAGA
 201 AAATGTACCC AAAACAGGGG AAACCTCAGT TACTAGTTGT TTTAAAAATG
 251 ACGCTGCAGC TGGAGATCTA AATTCTCTAG GAGGGGGATT TTCTTTCACA
 301 TTTAGCAATA TCGATGCAAC CACGGCTTCT GGAGCTGCTA TTGGAAGTGA
 351 AGCAGCTAAT AAGACAGTCA CGTTATCAGG ATTTTCGGCA CTTTCTTTTC
 401 TTTAAATCCCC AGCAAGTACA GTGACTAATG GATTGGGAGC TATCAATGTT
 451 AAAGGGAATT TAAGCCTATT GGATAATGAT AAGGTATTGA TTCAGGACAA
 501 TTTCTCAACA GGAGATGGCG GAGCAATTAA TTGTGCAGGC TCCTTGAAGA
 551 TCGCAAACAA TAAGTCCCTT TCTTTTATTG GAAATAGTTC TTCAACACGT
 601 GGCGGAGCGA TPCATACCAA AAACCTCACA CTATCTTCTG GTGGGGAAC
 651 TCTATTTTCA GGAATACAGC CGCTACGGC TGCTGGTAAA GGAGGTGCTA
 701 TCGCGATTGC AGACTCTGGC ACCCTATCCA TTTCTGGAGA CAGTGGCGAC
 751 ATTATCTTTG AAGGCAATAC GATAGGAGCT ACAGGAACCG TCTCTCATAG
 801 TGCTATTGAT TTAGGAACCTA GCGCTAAGAT AACTGCGTTA CGTGCTGCGC
 851 AAGGACATAC GATATACCTT TATGATCCGA TTAGTGTAA CAGTGGCGAC
 901 TCTGTTGCTG ATGCTCTCAA TATTAATAGC CCTGATACTG GAGATAACAA
 951 AGAGTATACG GGAACCATAG TCTTTTCTGG AGAGAAGCTC ACGGAGGCAG
 1001 AAGCTAAAGA TGAGAAGAAC CGCACTTCTA AATTACTTCA AAATGTTGCT
 1051 TTTAAAAATG GGACTGTAGT TTTAAAAGGT GATGTCGTTT TAAGTGGCAA
 1101 CGGTTTCTCT CAGGATGCAA ACTCTAAGTT GATTATGGAT TTAGGGACGT
 1151 CGTTGGTTGC AAACACCGAA AGTATCGAGT TAACGAATTT GGAATTAAT
 1201 ATAGACTCTC TCAGGAACGG GAAAAAGATA AAACCTCAGT CTGCCACAGC

-111-

5
10
15
20
25

```

1251 TCAGAAAGAT ATTCGTATAG ATCGTCCTGT TGTACTGGCA ATTAGCGATG
1301 AGAGTTTFTA TCAAAATGGC TTTTGAATG AGGACCATTC CTATGATGGG
1351 ATTCTTGAGT TAGATGCTGG GAAAGACATC GTGATTTCTG CAGATTCTCG
1401 CAGTATAGAT GCTGTACAAT CTCCGTATGG CTATCAGGGA AAGTGGACGA
1451 TCAATTGGTC TACTGATGAT AAGAAAGCTA CGGTTTCTTG GCGGAAGCAG
1501 AGTTTAAATC CCACTGCTGA GCAGGAGGCT CCGTTAGTTC CTAATCTTCT
1551 TTGGGGTTCT TTTATAGATG TTCGTTCCCT CCAGAATTTT ATAGAGCTAG
1601 GTACTGAAGG TGCTCCTTAC GAAAAGAGAT TTTGGGTTGC AGGCATTTCC
1651 AATGTTTTCG ATAGGAGCGG TCGTGAAAAT CAAAGGAAAT TCCGTCATGT
1701 GAGTGGAGGT GCTGTAGTAG GTGCTAGCAC GAGGATGCCG GGTGGTGATA
1751 CCTTGCTCTT GGGTTTGTCT CAGCTCTTTG CGCGTGACAA AGACTACTTT
1801 ATGAATACCA ATTTGCAAAA GACCTACGCA GGATCTTTAC GTTTCAGCA
1851 CGATGCTTCC CTATACTCTG TGGTGAGTAT CCTTTTAGGA GAGGGAGGAC
1901 TCCGCGAGAT CCTGTTGCCT TATGTTTCCA AGACTCTGCC GTGCTCTTTC
1951 TATGGGCAGC TTAGCTACGG CCATACGGAT CATCGCATGA AGACCGAGTC
2001 TCTACCCCCC CCCCCCCGA CGCTCTCGAC GGATCATACT TCTTGGGGAG
2051 GATATGTCTG GGCTGGAGAG CTGGGAAGTC GAGTTGCTGT TGAAATACC
2101 AGCGGCAGAG GATTTTCCA AGAGTACACT CCATTGTGTA AAGTCCAAGC
2151 TGTTTACGCT CGCCAAGATA GCTTTGTAGA ACTAGGAGCT ATCAGTCGTG
2201 ATTTTAGTGA TTCGCATCTT TATAACCTTG CGATTCCTCT TGGAAATCAAG
2251 TTAGAGAAAC GGTTCGAGA GCAATATTAT CATGTGTAG CGATGTATTC
2301 TCCAGATGTT TGTCGTAGTA ACCCCAAATG TACGACTACC CTACTTTCCA
2351 ACCAAGGGAG TTGGAAGACC AAAGGTTGCA ACTTAGCAAG ACAGGCTGGT
2401 ATTGTTTCAGG CCTCAGGTTT TCGATCTTTG GGAGCTGCAG CAGAGCTTTT
2451 CGGGAACCTT GGCCTTGAAT GCGGGGATC TTCTCGTAGC TATAATGTAG
2501 ATGCGGGTAG CAAATCAAA TTTTAG

```

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 70A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 70B).

The cp6270 protein was also identified in the 2D-PAGE experiment (Cpn0013).

These experiments show that cp6270 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 71

35 The following *C.pneumoniae* protein (PID 4376402) was expressed <SEQ ID 141; cp6402>:

40

```

1 MNVADLLSHL ETLSSKIFQ DYGPNGLQVG DPQTPVKKIA VAVTADLETI
51 KQAVAAEENV LIVHHGIFWK GMPYPITGMI HKRIQLLIEH NIQLIAYHLP
101 LDAHPTLGNW WRVALDLNWH DLKPFSSSLP YLGVQGSFSP IDIDSFIDLL
151 SQYYQAPLKG SALGGPSRVS SAALISGGAY RELSSAATSQ VDCFITGNFD
201 EPAWSTALES NINFLAFGHT ATEKVGPKSL AEHLKSEFPI STTFIDTANP
251 F*

```

The cp6402 nucleotide sequence <SEQ ID 142> is:

45
50
55

```

1 ATGAATGTTG CGGATCTCCT TTCTCATCTT GAGACTCTTC TCTCATCAAA
51 AATATTTTTCAG GATTATGGAC CCAACGGACT TCAAGTTGGA GATCCCCAAA
101 CTCCGGTAAA GAAAATCGCT GTTGCAAGTTA CCGCAGATCT AGAAACCATATA
151 AAACAAGCTG TTGCGGCCGA AGCAAACGTT CTCATTGTAC ACCACGGAAT
201 TTTTGGGAAA GGTATGCCCT ATCCTATTAC CGGCATGATC CATAAGCGCA
251 TCCAATTACT AATAGAACAC AATATCCAAC TCATTGCCTA CCACCTTCTT
301 TTGGATGCTC ACCCTACCTT AGGAAATAAC TGGAGAGTTG CCTTGGATCT
351 AAATTGCGAT GACTTGAAGC CCTTTGGTTC TTCCCTCCCT TATTTAGGAG
401 TGCAAGGCTC TTTCTCTCCT ATCGATATAG ATTCTTTCAT TGACCTGTTA
451 TCTCAATATT ACCAAGCTCC CCTAAAAGGA TCTGCCTTGG GCGGCCCTC
501 TAGAGTCTCC TCAGCAGCTC TGATCTCAGG AGGAGCTTAT AGAGAAGCTCT
551 CTTCGGCAGC CACGTCCTCA GTCGATTGCT TCATCACAGG AAATTTTGAT
601 GAACCTGCAT GGTTCGACAGC TCTAGAAAGC AATATCAACT TCCTAGCATT
651 TGGACATACA GCCACAGAAA AAGTAGGTCC AAAATCTCTT GCAGAGCATC

```

701 TAAAAAGCGA ATTTCCTATT TCCACAACCT TTATAGATAC GGCCAACCCC
751 TTCTAA

The PSORT algorithm predicts cytoplasmic (0.158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 71A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 71B) and for FACS analysis.

These experiments show that cp6402 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 72

10 The following *C.pneumoniae* protein (PID 4376520) was expressed <SEQ ID 143; cp6520>:

1 MKHYLSFSPS ADFFSKQGA ETQVLFGERV LVKGSTCYAY SOLFHNELLLW
51 KPYPGHSFRS TLVPCTPEFH IHPNVSVSV DAFLDPWGIP LPFGTLLHVN
101 SQNTVIFPKD ILNHMTIWG SGTPQCDPRH LRRLNYNFFA ELLIKDADLL
151 LNFFYVWGR SVHESLEKPG VDCSGFINIL YQAQGYNVPR NAADQYADCH
15 WISSFENLPS GGLIFLYPKE EKRISHVMLK QDSSTLIHAS GGGKKVEYFI
201 LEQDGKFLDS TYLFFRNNQR GRAFFGIPRK RKAFL*

The cp6520 nucleotide sequence <SEQ ID 144> is:

1 ATGAAACACT ACCATCATTT TTCTCCTTCT GCTGATTTT TCTCTAAACA
51 GGGTGCTATT GAAACTCAAG TCCTTTTGG AGAGCGCGTC TTAGTCAAAG
201 GGAGCACCTG CTATGCATAT TCCCAATTAT TCCACAATGA GCTGTATGG
151 AAGCCCTATC CAGGTCATAG CTTTCGTTCT ACCCTAGTCC CCTGCACTCC
201 TGAATTTTCA ATCCATCCAA ATGTTTCTGT GGTTCCTGTG GATGCATTTT
251 TAGATCCTTG GGGGATCCCT CTTCTTTTGA GAACCTTACT CCATGTGAAT
301 TCTCAAATA CCGTTATTTT CCCTAAGGAT ATTCCTCAATC ATATGAACAC
25 CATCTGGGGC TCCGGCACAC CTCAATGCGA TCCTAGACAT CTACGTCGTC
401 TAAATTATAA CTTCTTTGCT GAACCTTTTAA TTAAAGACGC AGACCTTTTA
451 CTGAACCTTC CCTATGTATG GGGAGGACGG TCTGTACACG AAAGCTCTGA
501 AAAGCCGGGT GTTGATTTGT CCGGATTAT CAATATCCTT TACCAGGCAC
551 AGGGATACAA CGTCCCTAGA AACGCTGCAG ATCAATATGC GGATTGTCAT
301 TGGATCTCTA GCTTTGAGAA CCTTCCTTCT GGTGGGTAA TATTTCTTTA
601 CCCTAAAGAA GAAAGCGTA TTTCTCATGT TATGTTGAAA CAGGATAGTT
651 CCACCTCAT TCATGCTTCT GGTGGAGGGA AAAAGTGA GTATTTTCAAT
701 TTAGAACAAAG ATGGGAAGTT TTTAGATTTCG ACTATCTAT TTTTGAAGAA
751 TAATCAGAGG GGACGGGCAT TTTTGGGAT CCCTAGAAAA AGAAAAGCCT
801 TTCTGTAA

The PSORT algorithm predicts cytoplasmic (0.265).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 72A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 72B) and for FACS analysis.

40 These experiments show that cp6520 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 73

The following *C.pneumoniae* protein (PID 4376567) was expressed <SEQ ID 145; cp6567>:

45 1 MTSPPIPFQSS GDASFLAEQ QQLPSTSESQ LVTQLLTMK HTQALSETVL
51 QQQRDLPTA SIILQVGGAP TGGAGAPFQP GPADDHHPI PPPVVAQIE
101 TEITIRSEL QLMRSTLQQS TKGARTGVLV VTAILMTISL LAIIIIILAV
151 LGFTGVLPQV ALLMQGETNL IWAMVSGSII CFIALIGTLG LILTNKNTPL

201 PAS*

The cp6567 nucleotide sequence <SEQ ID 146> is:

```

1  ATGACCTCAC  CGATCCCCTT  TCAGTCTAGT  GCGCATGCCT  CTTTCCTTGC
5  51  CGAGCAGCCA  CAGCAACTCC  CGTCTACTTC  TGAATCTCAG  CTAGTAACCTC
101 AATTGCTAAC  CATGATGAAG  CATACTCAAG  CATTATCCGA  AACGGTTCTT
151 CAACAACAAC  GCGATCGATT  ACCAACCGCA  TCTATTATCC  TTCAAGTAGG
201 AGGAGCTCCT  ACAGGAGGAG  CGGGTGCGCC  TTTTCAACCA  GGACCGGCAG
251 ATGATCATCA  TCATCCCATA  CCGCCGCTTG  TTGTACCAGC  TCAAAATAGAA
10  301 ACAGAAATCA  CCACTATAAG  ATCCGAGTTA  CAGCTCATGC  GATCTACTCT
351 ACAACAAAGC  ACAAAGGAG  CTCGTACAGG  AGTTCTAGTG  GTTACTGCAA
401 TCTTAATGAC  GATCTCCTTA  TTGGCTATTA  TTATCATAAT  ACTAGCTGTG
451 CTTGGATTTA  CGGGCGTCTT  GCCTCAAGTA  GCTTTATTGA  TGCAGGGTGA
501 AACAAATCTG  ATTTGGGCTA  TGGTGAGCGG  TTCTATTATT  TGCTTTATTG
15  551 CGCTAATTGG  AACTCTAGGA  TTAATTTTAA  CAAATAAGAA  CACGCTCTTA
601 CCGGCTTCTT  AA

```

The PSORT algorithm predicts inner membrane (0.694).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 73A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 73B) and for FACS analysis.

- 20 These experiments show that cp6567 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 74

The following *C.pneumoniae* protein (PID 4376576) was expressed <SEQ ID 147; cp6576>:

```

25 1  MLIMRNKVL  QISILALIOT  PLTLFSTKVK  KEGHVVDISI  TIITEGENAS
51  NKHPLPKLKT  RSGALFSQLD  FDEDLRLILAK  EYDSVEPKVE  FSEGKTNIAL
101 HLIAKPSIRN  IHISGNQVVP  EHKILKTLQI  YRNDLFEREK  FLKGLDDLRT
151 YLKRGYFAS  SVDYSLEHNQ  EKGHIDVLIK  INEGPCGKIK  QLTFSGISRS
201 EKSDIQEFIQ  TKQHSTTTSW  FTGAGLYHPD  IVEQDSLAI  NYLHNNGYAD
251 AIVNSHYDLD  DKGNIILYMD  IDRGSRYTLG  HVHIQGFVFL  PKRLIEKQSQ
30  301 VGPNDLYCPD  KIWDGAHKIK  QTYAKYGYIN  TNVDVLFIPH  ATRPIVDVTY
351 EVSEGSFYKV  GLIKITGNTH  TKSDVILHET  SLFPGDTFNR  LKLEDTEQRL
401 RNTGYFQSVS  VYTVRSQLDP  MGNADQYRDI  FVEVKETTTG  NLGLFLGFSS
451 LDNLFGGIEL  SESNFDLFGA  RNIFSKGFR  LRGGGEHLFL  KANFGDKVTD
501 YTLKWKPHF  LNTFPWILGIE  LDKSINRALS  KDYAVQTYGG  NVSTTYILNE
35  551 HLKYGLFYRG  SQTSLHEKRK  PLLGPNIDSN  KGFVSAAGVN  LNYDSVDSPR
601 TPTTGIRGGV  TFEVSGLGST  YHFTKLSLNS  SIYRKLTRKG  ILKIKGEAOF
651 IKPYSNTTAE  GVPVSRFFFL  GGETTVRGYK  SFIIGPKYSA  TEPQGLSSL
701 LISEEFQYPL  IRQPNISAFV  FLDSGFVGLQ  EYKISLKDRL  SSAGFGLRFD
751 VMNNVPVMLG  FGWPFPRPTET  LNKEKIDVSQ  RFFFALGGMF  *

```

- 40 A predicted signal peptide is highlighted.

The cp6576 nucleotide sequence <SEQ ID 148> is:

```

1  ATGCTCATCA  TGCGAATAAA  AGTTATCTTG  CAAATATCTA  TTCTAGCGTT
51  AATCCAAACC  CCTTTAACTT  TATTTTCTAC  TGAAAAAGTT  AAAGAAGGCC
101 ATGTGGTGGT  AGACTCTATC  ACAATCATAA  CGGAAGGAGA  AAATGCTTCA
45  151 AATAAACATC  CCTTACCCAA  ATTAAGAGACC  AGAAGTGGGG  CTCTTTTTC
201 TCAATTAGAT  TTTGATGAAG  ACTTGAGAA  TCTAGCTAAA  GAATACGACT
251 CTGTTGAGCC  TAAAGTAGAA  TTTTCTGAAG  GGAAAACTAA  CATAGCCCTT
301 CACCTAATAG  CTAACCCCTC  AATTCGAAAT  ATTCATATCT  CAGGAAATCA
351 AGTCGTTCCCT  GAACATAAAA  TTCTTAAAC  CCTACAAATT  TACCGTAATG
50  401 ATCTCTTTGA  ACGAGAAAAA  TTTCTTAAGG  GTCTTGATGA  TCTAAGAACG
451 TATTATCTCA  AGCGAGGATA  TTTTCGATCC  AGTGTAGACT  ACAGTCTGGA
501 ACACAATCAA  GAAAAAGGTC  ACATCGATGT  TTTAATTAAA  ATCAATGAAG
551 GTCCTTGCGG  GAAATTTAAA  CAGCTTACGT  TCTCAGGAAT  CTCTCGATCA
601 GAAAAATCAG  ATATCCAAGA  ATTTATTCAA  ACCAAGCAGC  ACTCTACAAC

```


5
10
15
20
25
30
35

```

651 TACAAGTTGG TTTACTGGAG CTGGACTCTA TCACCCAGAT ATTGTTGAAC
701 AAGATAGCTT GGCAATTACG AATTACCTAC ATAATAACGG GTACGCTGAT
751 GCTATAGTCA ACTCTCACTA TGACCTTGAC GACAAAGGGA ATATTCTTCT
801 TTACATGGAT ATTGATCGAG GGTCGCGATA TACCTTAGGA CACGTCCATA
851 TCCAAGGGTT TGAGGTTTTG CCAAAACGCC TTATAGAAAA GCAATCCCAA
901 GTCGGCCCCA ATGATCTTTA TTGCCCGGAT AAAATATGGG ATGGGGCTCA
951 TAAGATCAAA CAAACTTATG CAAAGTATGG CTACATCAAT ACCAATGTAG
1001 ACGTTCCTCT CATCCCTCAC GCAACCCGCC CTATTTATGA TGTAACCTAT
1051 GAGGTAAGTG AAGGGTCTCC TTATAAAGTT GGGTTAATTA AAATTACTGG
1101 GAATACCCAT ACAAATCTG ACGTTATTTT ACACGAAACC AGTCTCTTCC
1151 CAGGAGATAC ATTCAATCGC TTAAAGCTAG AAGATACTGA GCAACGTTTA
1201 AGAAATACAG GCTACTTCCA AAGCGTTAGT GTCTATACAG TTCGTTCTCA
1251 ACTTGATCCT ATGGGCAATG CGGATCAATA CCGAGATATT TTTGTAGAAG
1301 TCAAAGAAAC AACAACAGGA AACTTAGGCT TATTCTTAGG ATTTAGTTCT
1351 CTTGACAATC TTTTGGAGG AATTGAACTA TCTGAAAGTA ATTTTGATCT
1401 ATTTGGAGCT AGAAATATAT TTTCTAAAGG TTTTCGTTGT CTAAGAGGCG
1451 GTGGAGAACA TCTATTCTTA AAAGCCAAC TCGGGGACAA AGTCACAGAC
1501 TATACTTTGA AGTGGACCAA ACCTCATTTT CTAAACACTC CTTGGATTTT
1551 AGGAATTGAA TTAGATAAAT CAATTAACAG AGCATATATCT AAAGATTATG
1601 CTGTCCAAAC CTATGGCGGG AACGTGAGCA CAACGTATAT CTTGAACGAA
1651 CACCTGAAAT ACGGTCTATT TTATCGAGGA AGTCAAACGA GTTTACATGA
1701 AAAACGTAAG TTCCTCCTAG GGCCAAATAT AGACAGCAAT AAAGGATTTG
1751 TCTCTGCTGC AGGTGTCAAC TTGAATTACG ATTTCTGAGA TAGTCCTAGA
1801 ACTCCAAC TA CAGGATTTCG CCGGGGGGTG ACTTTTGAGG TTTCTGGTTT
1851 GGGAGGAACT TATCATTTTA CAAACTCTCT TTTAAACAGC TCTATCTATA
1901 GAAAACTTAC GCGTAAAGGT ATTTTGAAAA TCAAAGGGGA AGCTCAATTT
1951 ATTAACCCTT ATAGCAATAC TACAGCTGAA GGAGTTCTCTG TCAGTGAGCG
2001 CTTCTTCCTA GGTGGAGAGA CTACAGTTCG GGGATATAAA TCCTTTATTA
2051 TCGGTCCAAA ATACTCTGCT ACAGAACCCTC AGGGAGGACT CTCTTCGCTC
2101 CTTATTTTCAG AAGAGTTTCA ATACCCTCTC ATCAGACAAC CTAATATTAG
2151 TGCCTTTGTA TTCTTAGACT CAGGTTTGTG CCGTTTACAA GAGTATAAGA
2201 TTTCGTTAAA AGATCTACGT AGTAGTGCTG GATTGGTCT CCGCTTCGAT
2251 GTAATGAATA ATGTTCTCTGT TATGTTAGGA TTTGGTTGGC CCTCCGTCC
2301 AACCGAGACT TTGAATGGAG AAAAAATTGA TGTATCTCAG CGATTCTTCT
2351 TTGCTTTAGG GGCATGTTT TAA

```

The PSORT algorithm predicts outer membrane (0.7658).

The protein was expressed in *E. coli* and purified as GST-fusion (Figure 74A), his-tag and his-tag/GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 74B) and for FACS analysis (Figure 74C).

40 The cp6576 protein was also identified in the 2D-PAGE experiment (Cpn0300).

These experiments show that cp6576 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 75

The following *C. pneumoniae* protein (PID 4376607) was expressed <SEQ ID 149; cp6607>:

45
50

```

1 MNKRQKDKLK ICVIISTLIL VGIFARAPRG DTFKTFKSE EAIYSNQCN
51 EDMRKILCDA IEHADEEIFL RIYNLSEPKI QQSLTRQAQA KNKVITYYQK
101 FKIPQILKQA SNVTLVEQPP AGRKLMHQKA LSIDKKDAWL GSANYTNLSL
151 RLDNNLILGM HSSELCDLII TNSGDFSIIK DQTKYFVLP QDRKIAIQAV
201 LEKIQTAAQT IQVAMFALTH SEIIQALHQA KQGIHVDDII IDRSHSKLTF
251 QLRQLNINK DFVSINTAPC TLHHKFAVID NKTLLAGSIN WSKGRFSLND
301 ESLIILENLT KQONQKLRMI WKDLAKHSEH PTVDEEKEI IEKSLPVEEQ
351 EAA*

```

A predicted signal peptide is highlighted.

The cp6607 nucleotide sequence <SEQ ID 150> is:

1 ATGAATAAAA GACAAAAAGA TAAATTAAAA ATCTGTGTGA TTATTAGCAC
 51 GTTGATTTTA GTAGGAATTT TTGCAAGAGC TCCTCGTGGT GACACTTTTA
 101 AGACTTTTTT AAAGTCTGAA GAAGCTATCA TCTACTCAA TCAATGCAAT
 5 GAGGACATGC GTAAAATTCT ATGCGATGCT ATAGAACACG CTGATGAAGA
 201 GATCTTCCTA CGTATTTATA ACCTCTCAGA ACCCAAGATC CAACAGAGTT
 251 TAACTCGACA AGCTCAAGCA AAAAACAAAG TTACGATCTA CTATCAAAAA
 301 TTTAAAATTC CCCAAATCTT AAAGCAAGCC AGCAATGTAA CTTTAGTCGA
 351 GCAACCTCCA GCAGGGCGTA AACTGATGCA TCAAAAAGCT CTTTCCATAG
 401 ATAAGAAAGA TGCTTGGCTA GGATCTGCCA ACTACACCAA TCTTCTCTA
 10 CGTTTAGATA ATAATCTCAT TCTAGGAATG CATAGCTCGG AGCTCTGTGA
 501 TCTCATTTATC ACAAATACCT CTGGAGACTT TTCTATAAAG GATCAAAACG
 551 GAAAGTATTT TGTCTTCTCT CAAGATCGTA AAATTGCAAT ACAAGCTGTA
 601 CTCGAAAAAA TCCAGACAGC TCAGAAAACC ATCCAAGTTG CTATGTTTGC
 651 TCTGACCCAC TCGGAGATTA TTCAAGCCTT ACATCAAGCA AAACAACGAG
 15 GAATCCATGT AGATATTTATC ATTGATAGAA GTCATAGCAA ACTTACTTTT
 751 AAGCAATTAC GACAATTAAA TATCAATAAA GACTTTGTTT CTATAAATAC
 801 CGCACCTGT ACTCTTCACC ATAAGTTTGC AGTTATAGAT AATAAAACTC
 851 TACTTGACAG ATCTATAAAT TGGTCTAAAG GAAGATTCTC CTTAAATGAT
 901 GAAAGCTTGA TCATACTGGA AAACCTGACC AAACAACAAA ATCAGAAACT
 20 TCGAATGATT TGGAAAGATC TAGCTAAGCA TTCAGAACAT CCTACAGTAG
 951 ACGATGAAGA AAAAGAAATT ATAGAAAAAA GTCTTCCAGT AGAAGAGCAA
 1001 GAAGCAGCGT GA

The PSORT algorithm predicts periplasmic (0.934).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 75A) and also as a
 25 GST-fusion. The GST-fusion protein was used to immunise mice, whose sera were used in a Western
 blot (Figure 75B) and for FACS analysis.

These experiments show that cp6607 is a surface-exposed and immunoaccessible protein, and that it
 is a useful immunogen. These properties are not evident from the sequence alone.

Example 76

30 The following *C.pneumoniae* protein (PID 4376624) was expressed <SEQ ID 151; cp6624>:

1 MDAKMGYIFK VMRWIFCFVA CGITFGCTNS GFQNANSRPC ILSMNRMIHD
 51 CVERVVGNNRL ATAVLIKGS LDPHAYEMVK DKDKIAGSAV IFCNGLGLEH
 101 TSLRKHLEN NPNSVKLGER LIARGAFVPL EEDGICDPHI WMDLSIWKEA
 151 VIEITEVLIE KFPWESAIEFK ANSEELVCEM SILDSWAKQC LSTIPENLRY
 35 LVSCHNAFSY FTRRYLATPE EVASGAWRSR CISPEGLSPE AQISVRDIMA
 251 VVDYINEHDV SVVPEDTLN QDALKKIVSS LKKSHLVRLA QKPLYSDNVD
 301 DNYFSTFKHN VCLITEELGG VALECQR*

The cp6624 nucleotide sequence <SEQ ID 152> is:

1 ATGGATGCGA AAATGGGATA TATATTTAAA GTGATGCGTT GGATTTTCTG
 40 51 TTTCTGGGCA TGTGGTATAA CTTTGGATG TACCAATCTT GGGTTTCAGA
 101 ATGCAAATTC ACGTCCTTGT ATACTATCCA TGAATCGCAT GATTCATGAT
 151 TGTGTTGAAA GAGTCGTGGG GAATAGGCTT GCTACCGCTG TTTTGATCAA
 201 AGGATCCTTA GACCCTCATG CGTATGAGAT GGTAAAGGG GATAAGGACA
 251 AGATTGCTGG AAGTGCCGTA ATTTTGTGTA ACGGCTGGG TCTTGAGCAT
 45 301 ACATTAAGTT TCGGAAGCA TTTAGAAAAT AATCCCAATA GTGTCAAGTT
 351 AGGGGAGCGG TTGATAGCGC GTGGGCGCTT TGTCTCTCTA GAAGAAGACG
 401 GTATTGCGA TCCTCATATC TGGATGGATC TTTCTATTTG GAAGGAAGCT
 451 GTCATAGAAA TTACAGAAGT TCTCATTGAA AAGTTCCCTG AATGGTCTGC
 501 TGAATTTAAA GCAAATAGTG AGGAACCTGT TGTGAAATG TCTATTTTAG
 551 ATTCTTGGGC GAAACAATGC TTGAGCACAA TTCCTGAAAA TTTACGGTAT
 601 CTGTCTCAG GTCATAATGC GTTCAGTTAC TTTACACGTC GCTATTTAGC
 651 TACTCTGAA GAAGTGGCTT CCGGAGCATG GAGGTCTCGT TGTATTTCTC
 701 CTGAGGGTCT ATCTCCAGAA GCTCAAATCA GTGTTCTGTA TATTATGGCG
 751 GTTGATGATT ATATTAATGA GCATGATGTC AGTGTGGTTT TCCTGAGGA
 801 TACTCTGAAC CAAGATGCGT TGAAAAAAT TGTTCCTCTC CTGAAGAAAA
 851 GTCATTTAGT TCGTCTAGCT CAAAAACCAT TGTATAGTGA TAATGTGAGC
 901 GACAATTATT TTAGCACCTT TAAACATAAT GTCTGCCTTA TCACAGAAGA

951 ATTAGGAGGG GTGGCTCTTG AATGTCAAAG ATGA

The PSORT algorithm predicts inner membrane (0.168).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 76A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 76B) and for FACS analysis.

The cp6624 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6624 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 77

10 The following *C.pneumoniae* protein (PID 4376728) was expressed <SEQ ID 153; cp6728>:

```

1  MKSSVSWLFF SSIPLFSSLS IVAAEVTLDS SNNSYDGSNG TTFTVFSTTD
51  AAAGTTYSL SDVSFQNAGA LGIPLASGCF LEAGGDLTFQ GNQHALKFAF
101 INAGSSAGTV ASTSAADKNL LFNDFSRLSI ISCPSLLLSP TGQCALKSVG
151 NLSLTGNSQI IFTQNFSSDN GGVINTKNFL LSGTSQFASF SRNQAFATGKQ
201 GGVVYATGTI TIENSPGIVS FSQNLAKGSG GALYSTDNCS ITDNFQVIFD
251 GNSAWAQAQA QGGAICCTTT DKTVTLTGK NLSFTNTAL TYGGAISGLK
301 VSISAGGPTL FQSNISGSSA GQGGGGAINI ASAGELALSA TSGDITFNNN
351 QVTNGSTSTR NAINIIDTAK VTSIRAATGQ SIIFYDPITN PGTAASTDTL
401 NLNLADANSE IEYGGAI VFS GEKLSPT EKA IANVTSTIR QPAVLARGDL
20  451 VL RDGVTVTF KDLTQSPGSR ILM DGGT TLS AKEANLSLNG LAVNLSLSDG
501 TNKAALKTEA ADKNISLSGT IALIDTEGSF YENHNLSKAS TYPLLELT TA
551 GANGTITLGA LSTLTLEPE THYGYQGNWQ LSWANATSSK IGSINWTRTG
601 YIPSPERKSN LPLNSLWGNF IDIRSIQLI ETKSSGEFFE RELWLSGIAN
25  651 FFYRDSMPTR HGFRHISGGY ALGITATTPA EDQLTFAPCQ LFARDRNHIT
701 GKNHGDYGA SLYFHHT EGL FDIANFLWGK ATRAPWLSE ISQIIPLSFD
751 AKFSYLH TDN HMKTYTYD NS I IKG SWRND A F CADLGASLP FVISVPVLLK
801 EVEPFVKVQY IYAHQQDFYE RHAEGRAF NK SELINVEIPI GVTFERDSKS
851 EKGTYDLTLM YILDAYRRNP KCQTS LIASD ANWMAYGTNL ARQGF SVRAA
901 NHFQVNP HME IFQQFAFEVR SSSRNYN TNL GSKFCF*
```

30 The cp6728 nucleotide sequence <SEQ ID 154> is:

```

1  ATGAAGTCCT CTGTCTCTTG GTTGTCTTTT TCTTCAATCC CGCTCTTTTC
51  ATCGCTCTCT ATAGTCGCGG CAGAGGTGAC CTTAGATAGC AGCAATAATA
101 GCTATGATGG ATCTAACGGA ACTACCTTCA CGGTCTTTTC CACTACGGAC
151 GCTGCTGCAG GAAC TACCTA TTCTTACTTT TCCGACGTAT CTTTCAAAA
35  201 TGCAGGGGCT TTAGGAATTC CCTTAGCCTC AGGATGCTTC CTAGAAGCGG
251 GCGGCGATCT TACTTTCCAA GGAAATCAAC ATGCACTGAA GTTTGCATTT
301 ATCAATGCGG GCTCTAGCGC TGGAACTGTA GCCAGTACCT CAGCAGCAGA
351 TAAGAAATCTT CTCTTTAATG ATTTTCTTAG ACTCTCTATT ATCTCTTGTC
40  401 CCTCTCTTCT TCTCTCTCCT ACTGGACAAT GTGCTTTAAA ATCTGTGGGG
451 AATCTATCTC TAACTGGCAA TTCCCAAAT ATATTACTC AGAACTTCTC
501 GTCAGATAAC GCGGTGTGTA TCAATACGAA AAAC TCTTA TTATCAGGGA
551 CATCTCAGTT TGCGAGCTTT TCGAGAAACC AAGCCTTCAC AGGGAAGCAA
601 GCGGTGTAG TTTACGCTAC AGGAAC TATA ACTATCGAGA ACAGCCCTGG
45  651 GATAGTTTCC TTCTCTCAAA ACCTAGCGAA AGGATCTGGC GGTGCTCTGT
701 ACAGCACTGA CACTGTTCG ATTACAGATA ACTTCAAGT GATCTTTGAC
751 GGCAATAGTG CTTGGAAGC CGCTCAAGCT CAGGCGGGG CTATTTGTTG
801 CACTACGACA GATAAAACAG TGACTCTTAC TGGGAACAAA AACCTCTCTT
851 TCACAAATAA TACAGCATTG ACATATGGCG GAGCCATCTC TGGACTCAAG
901 GTCAGTATTT CCGCTGGAGG TCCTACTCTA TTTCAAAGTA ATATCTCAGG
50  951 AAGTAGCGCC GGTACGGGAG GAGGAGGAGC GATCAATATA GCATCTGCTG
1001 GGAAC TCGC TCTCTTGCT ACTTCTGGAG ATATTACCTT CAATAACAC
1051 CAAGTCACCA ACGGAAGCAC AAGTACAAGA AACGCAATAA ATATCATTGA
1101 TACCGCTAAA GTCACATCGA TACGAGCTGC TACGGGGCAA TCTATCTATT
1151 TCTATGATCC CATCAAAAT CCAGGAACCG CAGCTTCTAC CGACACATTG
55  1201 AACTTAACT TAGCAGATGC GAACAGTGAG ATCGAGTATG GGGGTGCGAT
1251 TGTCTTTTCT GGAGAAAAGC TTTCCCTTAC AGAAAAAGCA ATCGCTGCAA
```

1301 ACGTCACCTC TACTATCCGA CAACCTGCAG TATTAGCGCG GGGAGATCTT
 1351 GTACTTCGTG ATGGAGTCAC CGTAACCTTC AAGGATCTGA CTCAAAGTCC
 1401 AGGATCCCGC ATCTTAATGG ATGGGGGGAC TACACTTAGT GCTAAAGAGG
 1451 CAAATCTTTC GCTTAATGGC TTAGCAGTAA ATCTCTCCTC TTTAGATGGA
 5 1501 ACCAACAAGG CAGCTTTAAA AACAGAAGCT GCAGATAAAA ATATCAGCCT
 1551 ATCGGGAACG ATTGCGCTTA TTGACACGGA AGGGTCATTC TATGAGAATC
 1601 ATAACTTAAA AAGTGCTAGT ACCTATCCTC TTCTTGAAC TACCACCGCA
 1651 GGAGCCAACG GAACGATTAC TCTGGGAGCT CTTCTACCC TGACTCTTCA
 1701 AGAACCTGAA ACCCACTACG GGTATCAAGG AAACCTGGCAG TTGTCTTGGG
 10 1751 CAAATGCAAC ATCCTCAAAA ATAGGAAGCA TCAACTGGAC CCGTACAGGA
 1801 TACATTCCCTA GTCCTGAGAG AAAAAGTAAT CTCCCTCTAA ATAGCTTATG
 1851 GGGAAACTTT ATAGATATAC GCTCGATCAA TCAGCTTATA GAAACCAAGT
 1901 CCAGTGGGGA GCCTTTTGAG CGTGAGCTAT GGCTTTCAGG AATTGCGAAT
 1951 TTCTTCTATA GAGATTCTAT GCCCACCCGC CATGGTTTCC GCCATATCAG
 15 2001 CGGGGGTTAT GCACTAGGGA TCACAGCAAC AACTCCTGCC GAGGATCAGC
 2051 TTACTTTTGC CTTCTGCCAG CTCTTTGCTA GAGATCGCAA TCATATTACA
 2101 GGTAAGAACG ACGGAGATAC TTACGGTGCC TCTTTGTATT TCCACCATAC
 2151 AGAAGGGCTC TTCGACATCG CCAATTTCTC CTGGGGAAAA GCAACCCGAG
 2201 CTCCCTGGGT GCTCTCTGAG ATCTCCCGAG TCATTCTTTT ATCGTTCGAT
 2251 GCTAAATTCA GTTATCTCCA TACAGACAAC CACATGAAGA CATATTATAC
 20 2301 CGATAACTCT ATCATCAAGG GTTCTTGGAG AAACGATGCC TTCTGTGCAG
 2351 ATCTTGGAGC TAGCCTGCCT TTTGTTATTT CCGTTCGGTA TCTTCTGAAA
 2401 GAAGTCGAAC CTTTGTGCAA AGTACAGTAT ATCTATGCGC ATCAGCAAGA
 2451 CTTCTACGAG CGTCATGCTG AAGGACGCGC TTTCAATAAA AGCGAGCTTA
 25 2501 TCAACGTAGA GATTCCTATA GGCGTCACCT TCGAAAGAGA CTCAAAATCA
 2551 GAAAAGGGAA CTTACGATCT TACTCTTATG TATATACTCG ATGCTTACCG
 2601 ACGCAATCCT AAATGTCAAA CTTCCTTAAT AGCTAGCGAT GCTAAC TGGA
 2651 TGGCCTATGG TACCAACCTC GCACGACAAG GTTTTCTGT TCGTGCCTGC
 2701 AACCATTGCC AAGTGAACCC CCACATGGAA ATCTTCGGTC AATTGCTTT
 30 2751 TGAAGTACGA AGTTCTTAC GAAATTATAA TACAAACCTA GGCTCTAAGT
 2801 TTTGTTCTA G

The PSORT algorithm predicts inner membrane (0.187).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 77A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 77B) and for FACS analysis.

The cp6728 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6728 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 78

The following *C.pneumoniae* protein (PID 4376847) was expressed <SEQ ID 155; cp6847>:

1 MFVMMKKLVRL CVVLLSLLPN VLFSSDLLRE EGIKKMMDKL IEYHVDAQEV
 51 STDILSRSL S YTIQSFDPHK SYLSNQEVAV FLQSPETKKR LLKNYKAGNF
 101 AIYRNINQLI HESILRARQW RNEWVKNPKE LVLEASSYQI SKQPMQWSKS
 151 LDEVKQRQRA LLLSYLSLHL AGASSRYEG KEEQLAALCL RQIENHENVY
 45 201 LGINDHGVAM DRDEEAYQFH IRVVKALAH S LDAHTAYFSK DEALAMRIQL
 251 EKGMCIGVV LKEDIDGVVV REIIPGGFPA KSGDLQLGDI IYRVVDGKDIE
 301 HLSFRGVLD C LRGGHGSTVV LDIHRGESDH TIALRREKIL LEDRRVDVSY
 351 EPYGDGVIGK VTLHSFYEGE NQVSSEQDLR RAIQGLKEKN LLGLVLDIRE
 401 NTGGFLSQAI KVSGLFMTNG VVVVSRYADG TMKCYRTVSP KKFYDGPLAI
 50 451 LVSKSSASAA EIVAQTLQDY GVALVVGDEQ TYGKGTIQHQ TITGDASQDD
 501 CFKVTVGKYY SPSGKSTQLQ GVKSDILIPS LYAEDRLGER FLEHPLPADC
 551 CDNVLHDPLT DLDTQTRPWF QKYYPNLOK QETLWREMLP QLTKNSEQRL
 601 SENSNFQAF L SQIKSSEKTD LSYGSNDLQL BESINILKDM ILLQQCRK*

A predicted signal peptide is highlighted.

The cp6847 nucleotide sequence <SEQ ID 156> is:

1 ATGTTTCGTAA TGAAAAAACT TGTCGGTCTA TGCCTAGTTC TTCTTTCTTT
 51 ACTTCCGAAT GTATTATTTT CTTCGGATCT TTTACGAGAA GAGGGCATCA
 101 AAAAGATGAT GGACAAGCTG ATCGAGTATC ATGTCGATGC TCAAGAGGTT
 151 TCTACGGATA TACTCTCGCG TTCTTTATCT AGTTACATTC AATCTTTTGA
 201 TCCTCATAAA TCTTATCTTT CAAACCAAGA GGTTCAGTT TTTCTACAGT
 251 CTCCGGAAAC AAAGAAACGT CTCTTAAAGA ATTATAAGGC AGGCAACTTT
 301 GCTATTTATC GCAACATCAA TCAATTAATT CATGAGAGTA TTCTTCGTGC
 351 CAGGCAGTGG AGAAACGAAT GGGTTAAGAA TCCAAAAGAG CTTGTATTGG
 401 AGGCATCCTC ATATCAGATA TCGAAGCAAC CTATGCAATG GAGCAAATCT
 451 TTAGACGAAG TGAAGCAGAG ACAACGCGCT CTACTCCTTT CCTATCTTTT
 501 TTTACATCTT GCTGGAGCTT CTTCCTCTCG TTATGAGGGT AAAGAAGAGC
 551 AGCTTGCTGC TCTGTGTCTA CGTCAAATCG AGAACCATGA GAATGTATAT
 601 TTAGGTATCA ACGATCATGG TGTGTCTATG GATCGGGATG AAGAAGCCTA
 651 CCAATTCCAT ATCCGTGTTG TTAAAGCTTT AGCTCATAGC TTAGATGCAC
 701 ATACGGCGTA TTTCAGTAAG GACGAAGCGT TGGCGATGCG AATCCAACTA
 751 GAAAAAGGCA TGTGTGGAAT TGGTGTGTT CTGAAGGAAG ATATTGATGG
 801 AGTTGTTGTT AGAGAAATCA TTCCTGGGGG ACCTGCGGCT AAATCTGGGG
 851 ATCTTCAGCT TGGAGATATC ATCTATCGGG TGGATGGCAA GGATATCGAG
 901 CATCTTTCTT TCCGCGGTGT TTTAGATTGT TTACGTGGAG GTCATGGCTC
 951 TACTGTAGTC TTAGATATCC ATCGTGGGGA GAGCGATCAT ACGATCGCCT
 1001 TGAGAAAGGA GAAATCCTT TTAGAAGACC GTCGTGTGGA TGTTCCTAT
 1051 GAGCCTTATG GAGATGGTGT GATTGGGAAA GTTACGTTAC ATTCTTTTTA
 1101 TGAAGGAGAA AATCAGGTTT CTAGTGAACA AGATCTACGT CGAGCGATTC
 1151 AGGGATTAAA GGAGAAGAAC CTTCCTGGAT TAGTTTGA TAATCCGAGAA
 1201 AATACGGGTG GATTTTATC TCAAGCGATC AAAGTTCTG GTTTATTTAT
 1251 GACCAATGGC GTTGTGGTTG TATCTCGCTA TGCTGATGGT ACCATGAAGT
 1301 GCTACCGCAC AGTATCTCCT AAAAATTCT ATGATGGTCC TTTGGCTATT
 1351 TTAGTATCTA AAAGTTCCGC ATCAGCAGCG GAGATTGTAG CACAACTCT
 1401 CCAAGATTAT GGAGTTGCTT TAGTTGTTGG AGATGAGCAG ACCTATGGGA
 1451 AGGGAACGAT TCAGCATCAA ACAATTACTG GAGATGCCTC TCAGGACGAT
 1501 TGTTTTAAGG TTACTGTAGG GAAATATTAT TCCCTTCTG GGAATCGAC
 1551 TCAACTTCAG GGAGTAAAT CCGATATTTT AATTCCTTCT CTCTATGCTG
 1601 AAGATCGTCT AGGAGAGCGT TTTCTAGAGC ATCCCTTACC TGCAGATTGC
 1651 TGTGATAATG TACTTCACGA TCCTCTCAG GACTTGGATA CTCAAACACG
 1701 TCCTTGTTT CAAAATACT ATCTTCTAA TCTACAAAAG CAAGAGACTC
 1751 TTTGGAGAGA GATGCTACCT CAGCTTACGA AAAACAGTGA GCAAAGGCTT
 1801 TCTGAGAATT CGAATTTTCA GGCATTTTTC TCGCAGATAA AATCATCTGA
 1851 AAAAACGGAC CTATCCTATG GTTCCAATGA TTTACAATTG GAAGAGTCGA
 1901 TAAACATTTT GAAGGACATG ATTTTATTAC AACAGTGTAG AAAATAA

40 The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 78A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 78B) and for FACS analysis.

These experiments show that cp6847 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 79

The following *C.pneumoniae* protein (PID 4376969) was expressed <SEQ ID 157; cp6969>:

1 MRLFSLGTY LFFSLALSSC CGYSILNSPY HLSSLGKSL L QERIFIAPIK
 51 EDPHGQLCSA LTYELSKRSF AISGRSSCAG YTLKVELLNG IDKNIGFTYA
 101 PNKLGDKTHR HFIVSNEGR L SLSAKVQLIN NDTQEVLIQ CVARESVDFD
 151 FEPDLGTANA HEFALGQFEM HSEAIKSARR ILSIRLAETI AQQVYYDLF*

A predicted signal peptide is highlighted.

The cp6969 nucleotide sequence <SEQ ID 158> is:

1 ATGAGATTGT TTTCTTAGG CACGATTTAT CTTTTTTTTT CTCTAGCACT
 51 TTCGTCATGC TGTGGTTACT CTATTTTAAA CAGCCCGTAT CACTTATCGT
 101 CTTTAGGTAA GTCCTTATTA CAGGAAAGAA TTTTCATTGC TCCCATAAAA

5
151 GAAGATCCTC ATGGTCAGCT CTGCTCAGCT CTAACCTATG AGCTTAGTAA
201 GCGTTCCTTT GCTATCTCTG GAAGGAGTTC TTGCGCAGGC TATACTCTTA
251 AAGTAGAGCT TCTGAATGGT ATTGACAAGA ATATAGGTTT TACGTATGCC
301 CCAAATAAAC TCGGAGATAA GACTCACAGG CATTTTATAG TCTCTAATGA
351 AGGCAGACTA TCACTATCTG CAAAAGTACA GCTTATCAAT AATGACACTC
401 AAGAAGTCCT TATAGACCAA TGTGTTGCTC GAGAGTCTGT AGACTTTGAC
451 TTTGAGCCTG ACTTAGGAAC AGCAAACGCT CATGAATTG CTTTAGGCCA
501 ATTTGAAATG CATAGTGAAG CCATAAAAG TGCTCGCCGT ATACTATCTA
551 TACGCCTAGC CGAGACGATT GCTCAACAGG TATACTATGA CCTTTTGTGA

10 The PSORT algorithm predicts inner membrane (0.126).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 79A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 79B) and for FACS analysis.

15 These experiments show that cp6969 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 80

The following *C.pneumoniae* protein (PID 43771.09) was expressed <SEQ ID 159; cp7109>:

20
1 MKKTCQNYR SIGVVFSVVL FVLTTQTLEA GHFIDIGTSG LYSWARGVSG
51 DGRVVVGYEG GNAFKYVDGE KFLLEGLVPR SEALVFKASY DGSVIIGISD
101 QDPSCRAVKW VNGALVDLGI FSEGMSQFAE GVSSDGKTIV GCLYSDDTET
151 NFAVKWDETQ MVVLPNLPED RHSCAWDASE DGSVIVGDAM GSEELAKAVY
201 WKDGEQHLLS NIPGAKRSSA HAVSKDGSFI VGEFISEENE VHAFFVHNGV
251 IKDIGTLGGD YSVATGVSRD GKVIIVGHSTR TDGEYRAFKY VDGRMIDLGT
301 LGGSASFAGF VSDDGKTIVG KFETELGECH AFIYLLDD*

25 A predicted signal peptide is highlighted.

The cp7109 nucleotide sequence <SEQ ID 160> is:

30
1 ATGAAAAGA CATGTTGCCA AAATTACAGA TCGATAGGCG TTGTGTTCTC
51 TGTGGTACTT TTCGTTCTTA CAACACAGAC GCTGTTTGCA GGACATTTTA
101 TTGATATTGG AACTTCTGGA TTATATTCTT GGGCTCGAGG TGTATCTGGA
151 GATGGCCGCG TTGTCGTAGG TTATGAAGGT GGCAATGCAT TTAAATATGT
201 TGATGGTGAG AAATTCTCTG TAGAAGGTTT GGTCCCGAGA TCCGAGGCCT
251 TGGTATTTAA AGCTTCTTAT GATGGCTCTG TAATTATAGG AATCTCGGAT
301 CAAGATCCGT CTTGCCGCGC TGTGAAGTGG GTAAACGGTG CACTTGTGTA
351 TCTTGGAATA TTTTCTGAGG GAATGCAATC TTTTGCAGAG GGTGTTTCCA
401 GTGATGGAAG GACGATTGTA GGGTGCCTAT ATAGTGATGA TACAGAGACA
451 AACTTTGCTG TGAAGTGGGA TGAAACAGGA ATGGTTGTTC TCCCTAACTT
501 ACCAGAAGAT CGACATTCTT GCGCTTGGGA TGCCCTGTAA GATGGCTCTG
551 TGATTGTAGG GGACCCATG GGTAGCGAGG AAATTGCCAA GGCAGTGTAC
601 TGGAAGGACG GTGAACAACA TCTGCTTTCT AATATCCCAG GAGCTAAAAG
40 651 ATCGTCAGCA CATGCAGTTT CTAAAGATGG ATCTTTTATC GTAGGCGAGT
701 TCATCAGTGA AGAAAATGAA GTTCATGCCT TTGTTTATCA CAACGGTGTT
751 ATCAAAGATA TCGGGAATTT AGGAGGAGAT TACTCTGTAG CAACTGGAGT
801 TTCTAGGGAT GGTAAGGTCA TCGTGGGTCA TTCTACAAGA ACAGATGGTG
851 AATACCGTGC ATTTAAATAT GTGGATGGAA GAATGATAGA TTTGGGGACT
45 901 TTAGGAGGTT CAGCATCTTT TGCTTTTGGT GTTTCTGACG ATGGCAAAAC
951 AATCGTAGGA AAATTTGAAA CAGAGCTAGG AGAATGTCAT GCCTTTATCT
1001 ACCTTGATGA TTAG

The PSORT algorithm predicts outer membrane (0.887).

50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 80A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 80B) and for FACS analysis.

These experiments show that cp7109 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 81

The following *C.pneumoniae* protein (PID 4377110) was expressed <SEQ ID 161; cp7110>:

```

5      1  MAAIKQILRS  MLSQSSSLWMV  LFSLYSLSGY  CYVITDKPED  DFHSSSAVKW
      51  DHWGKTTLSR  LSNKKASAKA  VSGTGATTVG  FIKDTWSRTY  AVRWNWYWGK
     101  ELPTSSWVKK  SKATGISSDG  SIIAGIVENE  LSQSFAVTWK  NNEMYLLPST
     151  WAVQSKAYGI  SSDGSVIVGS  AKDAWSRTFA  VKWTGHEAQV  LPVGVAVKSV
     201  ANSVSANGSI  IVGSVQDASG  ILYAVKWEKN  TITHLGTLLG  YSAIAKAVSN
     251  NGKVIVGRSE  TTYGEVHAF  HKNGVMSDLG  TLGGSYSAAK  GVSATGKVIV
     301  GMSTTANGKL  HAFKYVGGRM  IDLGEYSWKE  ACANAVSIDG  EIIIVGVQSE*
  
```

A predicted signal peptide is highlighted.

The cp7110 nucleotide sequence <SEQ ID 162> is:

```

15      1  ATGGCAGCTA  TAAACAAAT  TTTACGTTCT  ATGCTATCTC  AGAGTAGCTT
      51  ATGGATGGTC  CTATTTTCAT  TATATTCTCT  ATCTGGTTAT  TGCTATGTAA
     101  TTACAGACAA  ACCAGAAGAT  GACTTCCATT  CTTCATCCGC  AGTAAATGG
     151  GATCATTGGG  GAAAGACAAC  TCTCTCAAGA  TTATCAAATA  AAAAGCCTC
     201  TGCAAAAGCT  GTTTCAGGAA  CTGGTGCTAC  AACTGTCCGC  TTTATAAAG
     251  ACACCTGGTC  TCGAACATAC  GCAGTAAGAT  GGAATTATTG  GGGGACCAAA
     301  GAACTCCCTA  CCAGTCATG  GGTAAAAAAA  TCAAAAGCAA  CAGGAATCTC
     351  CTCTGATGGG  TCTATAATCG  CGGGGATTGT  CGAGAATGAG  CTTTCTCAAA
     401  GTTTCGCAGT  CACATGGAAA  AACAAATGAAA  TGTATTTGCT  CCCTTCCACA
     451  TGGGCAGTGC  AATCTAAAGC  GTATGGAATT  TCTTCTGATG  GCTCTGTTAT
     501  TGTAGGGAGT  GCTAAGGATG  CTTGGTCGCG  AACTTTCGCT  GTGAAGTGGA
     551  CGGGACACGA  GGCTCAGGTG  TTACCAAGTAG  GCTGGGCTGT  CAAATCTGTA
     601  GCGAATTCTG  TATCTGCCAA  TGGATCTATA  ATTGTAGGGT  CTGTACAAGA
     651  CGCCTCTGGA  ATTCTTTATG  CTGTAAAGTG  GGAAGGGAAC  ACTATTACAC
     701  ATCTAGGAAC  TTTAGGAGGC  TATTCTGCCA  TTGCAAAAGC  TGTATCCAAT
     751  AATGGCAAGG  TCATTGTAGG  GAGATCCGAA  ACATATTATG  GAGAGGTCCA
     801  TGCTTTCTGT  CATAAGAATG  GCGTCATGTC  AGACCTCGGC  ACCCTCGGAG
     851  GATCTTATTC  TGCAGCTAAG  GGAGTCTCTG  CAACTGGAAA  AGTTATTGTC
     901  GGTATGTCCA  CAACAGCAAA  TGGGAAATTG  CATGCCCTTA  AATATGTCGG
     951  TGGAAGAATG  ATCGACTTAG  GAGAGTATAG  CTGGAAAGAA  GCCTGTGCAA
    1001  ACGCTGTTTC  TATTGATGGA  GAAATTATTG  TTGGAGTCCA  ATCAGAATAA
  
```

35 The PSORT algorithm predicts outer membrane (0.827).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 81A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 81B) and for FACS analysis.

40 These experiments show that cp7110 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Figure 191 shows a schematic representation of the structural relationships between of cp7105, cp7106, cp7107, cp7108, cp7109 and cp7110, each of which is identified herein. These six proteins may be grouped in a new family of related outer membrane-associated proteins. These proteins have a repeat structure in common (cf. the pmp family).

45 Example 82

The following *C.pneumoniae* protein (PID 4377127) was expressed <SEQ ID 163; cp7127>:

```

1  MVFFRNSLLH  LVALSGMLCC  SSGVALTIAE  KMASLEHSGR  GADDYEGMAS
  
```


51 FNANMREYSL QLSKLYEEAR KLRASGTEDE ALWKDLIRRI GEVRGYLREI
 101 EELWAAEIRE KGGNLEDYAL WNHPEITTYN LVTDVGTEDS IYLIPOEIGA
 151 IKIATLSKFV VPKESEFEDCL TQILSRLGIG VRQVNSWIKE LYMMRKEGCS
 201 VAGVFSSRKD LEALPETAYI GFVLNSNVDA HTNQHVLLKF INPETHVDV
 251 IAGRVWIFGS AGEVGELLKI YNFVQSESIR QEYRVIPLTK IDPGEMISIL
 301 NAAFREDLTK DVSEESLGLR VVPLQYQGRS LFLSGTAALV QQALTLIREL
 351 BEGIENPTDK TVFWYNVKHS DPQELAALLS QVHDFVSGEN KASVGAADGC
 401 GSQNASIQI DTTVSSSAKD GSVKYGNFIA DSKTGTLMV VEKLEVLPRIO
 451 MLLKKLDVPG KMVRIEVLLE ERKLAHEQKS GLNLLRLGEE VCKKGCSPSV
 501 SWAGGTGILE FLFKGSTGSS IVPGYDLAYQ FLMAQEDVRI NASPSVVTMN
 551 QTPARIAVD EMSIAVSSDK DKAQYNRAQY GIMIKMLPVI NVGEEDGKSY
 601 ITLETDTITFD TTGNHDDRFP DVTRRNITNK VRIADGETVI IGGLRCKQMS
 651 DSHDGIPFLG DIPGIGKLF MSSTSDSLTE MFVFTPKIL ENPVEQQRK
 701 EEALLSSRPG EREYYQALA ASEAAAAAH KKLEMFPSG VLSQVERQE
 751 YDGC*

A predicted signal peptide is highlighted.

The cp7127 nucleotide sequence <SEQ ID 164> is:

1 ATGGTTTTTT TCCGTAATTC TTTACTGCAT TTAGTTGCC TATCCGGAAT
 51 GCTCTGTGT TCTTCTGGAG TGGCTTTAAC GATAGCCGAG AAGATGGCTT
 101 CTTTAGAGCA CTCGGGGAGA GGAGCAGACG ATTATGAGGG GATGGCTTCG
 151 TTTAATGCCA ATAGAGGGA GTATAGCCTT CAGCTGAGCA AGTTGTATGA
 201 GGAAGCACGA AAGCTACGCG CTTCTGGAAC TGAGGATGAA GCTCTGTGGA
 251 AGGACTTAAT TCGACGGATT GGTGAGGTGC GAGGCTATCT TCGAGAGATC
 301 GAGGAGCTTT GGGCTGCAGA AATTCGTGAG AAAGGGGGCA ATCTCGAGGA
 351 CTACGCCCTC TGGAATCACC CAGAGACTAC GATTTACAAT CTGTATTACC
 401 ATTACGGAAC CGAAGACTCT ATTTATTGTA TTCCTCAAGA AATCGGAGCG
 451 ATTAAATCG CAACCTTATC GAAATTTGTA GTTCCTAAAG AGTCTTTCGA
 501 AGACTGTCTC ACTCAGATCC TATCTCGCTT AGGTATTGGC GTGCGTCAGG
 551 TCAATTCTTG GATTAAGGAA CTTTATATGA TCGCTAAGGA GGGCTGCAGT
 601 GTTGCTGGAG TTTTTCCTC CAGAAAAGAT TTAGAGGCGC TCCAGAAAC
 651 AGCCTATATT GGTTTGTAT TGAATTCGAA CGTAGATGCG CATACCAATC
 701 AACATGTCTT AAAAAAGTTC ATTAACCTTG AAACAACGCA TGTAGATGTG
 751 ATTGCAGGAC GTGTGTGGAT TTTTGGTTCT GCGGGGGAAG TCGGCAGACT
 801 TCTGAAGATT TATAATTTTG TGCAGTCGGA GAGCATACGT CAAGAGTATC
 851 GGGTGATTCC CTTAACTAAG ATCGATCCAG GGGAGATGAT TTCCATTCTC
 901 AACGCAGCAT TTCGTGAGGA TCTGACTAAA GATGTTAGTG AAGAACTCTT
 951 AGGCCCTTCGT GTAGTTCCCTT TACAGTATCA AGGGCGTTCT TGTGTTTTAA
 1001 GTGGAACCGC GCGTTAGTG CAGCAAGCGC TGACTCTCAT TCGAGAGCTT
 1051 GAAGAAGGGA TTGAGAAGCC TACGGATAAA ACAGTATTTT GGTATTAACGT
 1101 CAAGCACTCC GATCCCAAG AGTTGGCGGC ATTGCTTTCC CAAGTCCATG
 1151 ATGCTCTCTC TGGCGAGAAAT AAGGCGAGTG TCGGAGCTGC AGATGGATGT
 1201 GGGTCGCAAT TAAATGCCTC GATCCAAATT GATACTACAG TAAGTTCTTC
 1251 TGCGAAAGAT GGCTCAGTGA AGTACGGAAT CTTTCATCGCG GATTCTAAGA
 1301 CAGGAACCTC GATTATGGTG GTTGAGAAAG AAGTTCTTCC ACGTATTTCAG
 1351 ATGCTACTTA AGAACTAGA TGTCCCTAAA AAGATGGTCC GTATCGAGGT
 1401 GCTGTATTTT GAAAGAAAT TGGCACATGA GCAGAAATCT GGGTTAAATC
 1451 TTCTACGTCT TGGTGAGGAA GTTTGTAAAA AAGGGTGCAG TCCTTCTGTG
 1501 TCTTGGGCGG GGGGTACTGG CATACTAGAA TTTTATTTA AAGGAAGTAC
 1551 GGGATCTTCG ATAGTTCTCG GTTATGATCT CGCCTATCAA TTTTAAATGG
 1601 CTCAGAGGGA CGTTCGGATT AATGCGAGTC CTTCTGTAGT TACTATGAAC
 1651 CAAACCCAG CACGGATTGC TGTGTTGAT GAAATGTCAA TAGCGGTGTC
 1701 TTCAGATAAA GATAAAGCGC AATACAATCG TGCGCAGTAC GGTATCATGA
 1751 TAAAAATGCT CCCCGTAAT AATGTGGGAG AGGAAGACGG AAAAGATTAC
 1801 ATTACTTTAG AGACAGACAT CACCTTTGAT ACTACGGGAA AAAATCATGA
 1851 TGATCGTCTT GATGTTACAA GCGTAATAT TACTAATAAG GTGCGCATTG
 1901 CTGACGGAGA GACTGTGATT ATTGGAGGTT TGCGTTGCAA ACAGATGTCA
 1951 GATTCTCATG ATGGCATTC TTTCTTGGA GACATTCCTG GTATAGGGAA
 2001 GTTATTTGGA ATGAGTTCCA CATCAGACAG TCTCAGGAG ATGTTTGTAT
 2051 TTATCACTCC GAAGATCCTA GAAAATCCTG TAGAGCAACA AGAACGTAAA
 2101 GAAGAAGCTT TACTCTCTTC GCGCCCTGGA GAGAGAGAAG AATACTATCA
 2151 GGCCTTAGCA GCTAGTAGG CTGCAGCACG AGCAGCTCAT AAAAATTAG
 2201 AGATGTTCCT GGCATCAGGA GTATCTTTAT CTGAGGTAGA GAGGCAAGAA
 2251 TACGATGGCT GCTAG

The PSORT algorithm predicts periplasmic (0.920).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 82A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 82B) and for FACS analysis.

These experiments show that cp7127 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 83

The following *C.pneumoniae* protein (PID 4377133) was expressed <SEQ ID 165; cp7133>:

```

1  MQPFIFITLLC LTSLVSLVAF DAANARKRCA CAQTIERGEN FFSIKRSACA
51  EIEYQEKSRH ASAIERISKD KGKVTPKQIA KVATKKKQRY RLLQVPFSRP
101 PNNRYNLYA LLSEPPPECYS DTASWYAIFI RLLRRAYVDI GNVPPGSEYA
151 IANALISNKQ EILERGAQLG PDVIETLTLP EEQAEIFYKM LKGSNSQSLS
201 LNFLHYEEKS LGHCKLNLIF MDPLLEAVL DHPDAYRETS LLRDGIWEAV
251 KRQEHAIQEH GQAAALELFK TRTDFRLELR DKMQLLSRY DLLPLLNKKM
301 FDYTLGSAGD YLFLVDPDTK AISRCRCPSK SIKL

```

A predicted signal peptide is highlighted.

The cp7133 nucleotide sequence <SEQ ID 166> is:

```

1  ATGCAACCTT TTATCTTTAC TTTACTGTGC TTGACATCTT TGGTTTCTTT
51  AGTCGCCCTT GATGCTGCGA ATGCTCGTAA ACGTTGTGCC TGTGCTCAAA
101 CTATAGAACG TGGAGAGAAC TTCTTTTCCA TAAACGCTC TGCTTGTGCT
151 GAAATCGAAT ATCAAGAAAA ATCTCGCCAC GCCTCAGCAA TTGAAAGAA
201 CTCAAAGAT AAAGGCAAAG TCACTCCAAA GCAGATTGCG AAAGTAGCTA
251 CTAAGAAAAA GCAAAGATAC CGTTTATTGC AGGTTCCCTT TTCAAGGCCT
301 CCGAATAACT CAAGGTATAA CCTCTATGCT TTGCTTAGTG AACCTCCCGA
351 ATGCTATAGC GATACAGCAT CATGGTATGC TATTTTATT CGGTTACTTC
401 GACGTGCTTA TGTAGACACG GGAAATGTAC CTCCTGGATC TGAGTATGCC
451 ATCGCTAATG CTTTGATAAG TAACAAACAA GAGATTTTAG AGAGGGGAGC
501 GCAGCTTGA CCGATGTTA TTGAAACTCT AACATTGCCT GAGGAACAA
551 CCGAGATTTT TTATAAAATG CTCAAAGGGT CGTCAAAC TCAGTCGCTA
601 CTGAATTTTC TGCATTATGA AGAGAAAAGC TTAGGCCACT GTAAGCTAAA
651 TCTGATCTTC ATGGATCCCC TACTGTTAGA AGCTGTTCTA GATCATCCCG
701 ATGCTTATAG GGAAACGTCG CTCCTGCGCG ATGGCATTTG GGAAGCGGTG
751 AAGCGTCAAG AACATGCCAT CCAAGAACAT GGCCAGGCAG CTGCTTTTGA
801 GCTTTTAAA ACACGCACCG ACTTCCGCTT GGAGCTGCGA GATAAGATCC
851 AGTTACTTCT AAGTCGATAC GATTGTCTCC CTTATATAA TAAAAAATG
901 TTCGACTACA CTTAGGAAG TGCCGGAGAT TACTATTTT TGGTAGACCC
951 AGATACTAAG GCAATTTCTC GATGTCGCTG CCCTCAAAG AGTATTAAAT
1001 TATAA

```

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 83A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 83B) and for FACS analysis.

These experiments show that cp7133 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 84

The following *C.pneumoniae* protein (PID 4377222) was expressed <SEQ ID 167; cp7222>:

```

1  MNRRDMVITA VVVNAILLVA LFVTSKRIGV KDYDEGFRNF ASSKVTOAVV
51  SEEKVIEKPV VAEVPSRPIA KETLAAQFIE SKPVIITTPP VPVVSETPEV

```

101 PTVAVPPQPV RETVKEEQAP YATVVVKKGD FLERIANRH TTVAKLMQIN
 151 DLTTTQLKIG QVIKVPSTQD VSNEKTPQTQ TANPENYYIV QEGDSPWTIA
 201 LRNHIRLDDL LKMNDLDEYK ARRLKPGDQL RIR*

A predicted signal peptide is highlighted.

- 5 The cp7222 nucleotide sequence <SEQ ID 168> is:

1 ATGAATCGTA GAGACATGGT AATAACAGCT GTCCTAGTGA ATGCTATATT
 51 GCTTGTGGCT CTTTTCGTCA CATCAAAGCG TATTGGCGTC AAGGACTATG
 101 ACGAGGGATT CCGTAATTTT GCTTCTAGCA AGGTTACACA AGCAGTAGTT
 151 TCAGAAGAAA AAGTCATAGA AAAGCCTGTA GTCGCAGAAG TGCCTAGCCG
 10 TCCTATCGCT AAAGAGACTC TAGCTGCACA GTTTATTGAA AGTAAGCCGG
 201 TATTTGTAAC CACACCACCC GTGCCTGTTG TTAGCGAAAC CCCAGAAGTG
 251 CCTACTGTGG CAGTTCCGCC TCAGCCTGTT CGTGAGACAG TAAAAGAGGA
 301 ACAAGCTCCT TATGCTACTG TTGTAGTGAA AAAAGGAGAT TTCTCGAAC
 351 GCATTGCGAG AGCAAATCAT ACTACCGTTG CAAAATTGAT GCAGATCAAT
 15 GATCTTACCA CCACCAACT TAAAATTGGT CAGGTCATCA AAGTCCCTAC
 501 GTCTCAAGAT GTCAGCAACG AAAAACTCC TCAAACACAG ACCGCAAACC
 551 CTGAAAATTA TTATATCGTC CAAGAAGGGG ATAGCCCCGTG GACAATAGCA
 601 TTGCGTAACC ATATTCCGAT GGATGATTG CTAAAAATGA ATGATCTCGA
 651 TGAATATAAA GCCCGGCGCC TTAAGCCTGG AGATCAGTTG CGCATACGTT
 20 GA

The PSORT algorithm predicts periplasmic (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 84A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 84B) and for FACS analysis.

- 25 These experiments show that cp7222 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 85

The following *C.pneumoniae* protein (PID 4377225) was expressed <SEQ ID 169; cp7225>:

1 MKGTPQYHFI GIGGIGMSAL AHILLDRGYE VSGSDLYESY TIESLKAKGA
 30 51 RCFSGHDSH VPHDAVVVYS SSIAPDNVEY LTATQRSSRL LHRAELLSQL
 101 MEGYESILVS GSHGKTGTSS LIRAIHQEAQ KDPSYAIGGL AANCLNGYS
 151 SSKIFVAEAD ESDGSLKHVT PRAVVITNID NEHLNNYAGN LDNLVQVIQD
 201 FSRKVTDLNK VFYNGDCPIL KGNVQGISYG YSPECQLHIV SYNQKAWQSH
 35 FSFTFLGQY QDIELNLPQ HNAANAAAAC GVALTFGIDI NIIRKALKKF
 301 SGVHRRLEK NISESFLFLE DYAHHPVEVA HTLRSVRDAV GLRRVIAIFQ
 351 PHRFSLREC LQTFPKAFQE ADEVILTDVY SAGESPREST ILSDLAEQIR
 401 KSSVYHCCYV PHGDIVDYLR NYIRIHVVCV SLGAGNIYTI GEALKDFNPK
 451 KLSIGLVCQG KSCHEHDSLL SAQHVSKYIS PEFYDVSYFI INRQGLWRTG
 501 KDFPHLIEET QGDSPLSSEI ASALAKVDCL FVVLHGPFGE DGTIQGFPEI
 40 LGKPYAGPSL SLAATAMDKL LTKRIASAVG VPVVPYQPLN LCFWKRNP
 551 CIQNLLETFS FPMIVKTAHL GSSIGIFLVR DKEELQEKIS EAFLYDTDV
 601 VESRLGSR EIVSCIGHSS SWYCMAGPNE RCGASGFIDY QEKYGFDDID
 651 CAKISFDLQL SQESLDCVRE LAERVYRAMQ GKGSARIDFF LDEEGNYWLS
 701 EVNPIPGMTA ASPFLQAFVH AGWTQEQIVD HFIIDALHKF DKQQTIEQAF
 45 801 TKEQDLVKR*

The cp7225 nucleotide sequence <SEQ ID 170> is:

1 ATGAAGGGA CTCCTCAGTA TCATTTTATC GGTATCGGTG GTATAGGAAT
 51 GAGCGCTTTA GCTCATATTT TGCTTGATCG TGGCTATGAG GTCTCTGGAA
 101 GCGACTTATA TGAAAGCTAT ACGATCGAAA GCCTGAAAGC TAAAGGTGCG
 50 151 AGGTGTTTCT CAGGCCATGA TTCTCCCAT GTTCTCATG ATGCCGTGCT
 201 TGTTTATAGC TCAAGTATAG CCCCTGATAA GTAGAGTAT CTTACCGCTA
 251 TTCAAAGATC ATCAGTCTT CTTCATAGAG CAGAGCTCTT GAGTCAGCTT
 301 ATGGAGGGTT ATGAAAGCAT TCTGGTTTCA GGAAGCCATG GGAAGACAGG
 351 GACCTCATCT CTAATTCGAG CGATTTTCCA GGAAGCTCAG AAAGATCCCT

5
 10
 15
 20
 25
 30
 35
 40

```

401 CCTATGCTAT TGGAGGACTC GCTGCAAACT GCCTGAATGG GTATTCTGGA
451 TCATCGAAAA TCTTCGTTGC CGAAGCCGAT GAAAGTGATG GGTCTTTTAAA
501 GCACTACACT CCCCGTGCAG TAGTCATTAC AAATATAGAT AATGAACATT
551 TGAATAATTA CGCTGGGAAT CTTGATAACC TGGTTCAGGT AATCCAGGAC
601 TTCTCTAGAA AAGTAACAGA TCTCAATAAG GTATTCTATA ACGGGGATTG
651 TCCTATTTTG AAAGGAAATG TCCAAGGGAT TTCTTATGGA TATTCAACAG
701 AATGTCAATT GCATATCGTT TCCTATAATC AAAAGGCATG GCAATCTCAC
751 TTTTCCTTFA CCTTTTTAGG CCAGGAGTAT CAAGACATG AGCTCAATCT
801 CCCTGGACAA CATAACGCTG CAAATGCAGC AGCAGCCTGT GGAGTTGCTC
851 TTACCTTTGG CATAGACATA AACATCATTC GAAAAGCTCT CAAAAAATTC
901 TCGGGAGTTC ATCGACGCTT AGAAAGAAAA AATATATCCG AAAGCTTTCT
951 TTCTCTAGAA GATTATGCTC ATCATCCTGT AGAGGTTGCA CATACCCTGC
1001 GCTCTGTGCG TGATGCTGTG GGTTCGCGAA GAGTCATCGC AATTTTTCAA
1051 CCACATCGAT TCTCTCGTTT AGAAGAGTGC TTACAAACCT TCCCCAAAGC
1101 TTCCAAGAA GCTGATGAAG TCATACTTAC AGATGTCTAT AGTGCCGGAG
1151 AAAGTCCTAG AGAGTCTATC ATCTTTTCCG ACCTTGCGGA ACAGATTTCGT
1201 AAGTCTTCTT ATGTCCATTG TTGTTATGTT CCCCATGGAG ACATCGTAGA
1251 TTATCTACGA AACTACATTC GCATTCTATG TGTCTGTGTT TCTCTAGGAG
1301 CTGGAAATAT CTATACTATT GGAGAGGCTT TAAAGACTT TAACCCATAA
1351 AAATTATCCA TAGGACTCGT CTGTGGAGGG AAATCTTGCG AACACGATAT
1401 TTCTCTACTT TCTGCTCAAC ATGTCTCTAA ATATATTTCT CCTGAATTCT
1451 ATGATGTGAG TTACTTCATC ATAAATCGTC AGGGCTTATG GAGAACAGGA
1501 AAGGATTTTC CTCATCTTAT TGAAGAGACT CAAGGGGATT CGCCACTTTC
1551 TTCTGAAATC GCTTCAGCTT TAGCAAAAGT CGACTGTTTG TTTCCCGTGC
1601 TCCATGGCCC ATTTGGAGAG GATGGTACGA TCCAGGGATT TTTTGAATC
1651 TTAGGAAAAC CTTATGCCGG ACCCTCACTA TCTTTAGCAG CAACTGCAAT
1701 GGATAAGCTG TTAACAAAAC GAATTGCATC AGCAGTGGGT GTTCTGTAG
1751 TCCCTTACCA ACCTTTAAAT CTCTGTTTCT GGAAACGCAA TCCAGAACTA
1801 TGTATTCAGA ATCTTATAGA GACATTTTCT TTCCCTATGA TTGTAAAAC
1851 TGCACATTG GATCTAGTA TTGGGATATT TTTAGTCCGT GATAAAGAGG
1901 AATTACAAGA AAAGATCTCA GAAGCATTC TATATGACAC GGTGTGTTT
1951 GTGGAGGAAA GTCGCTTAGG GTCTCGTGAA ATCGAAGTGT CCTGTATCGG
2001 CCATPCTTCT AGCTGGTATT GTATGGCAGG GCCTAATGAA CGCTGTGGTG
2051 CTAGTGGGTT TATTGATTAT CAAGAGAAAT ATGGATTTGA TGGCATAGAT
2101 TCGCAAAGA TCTCTTTTGA TTTACAGCTC TCACAAGAAT CTTTAGATTG
2151 TGTTAGAGAA CTTGCAGAGC GTGTCTACCG AGCAATGCAA GGAAAAGGTT
2201 CAGCTCGAAT AGATTTTTTTC TTGGATGAAG AGGGGAATTA TTGGTTGTCA
2251 GAGGTCAATC CTATTCCAGG AATGACAGCA GCTAGCCCAT TTTTACAAGC
2301 TTTTGTTCAC GCAGGATGGA CGCAAGAACA AATTGTAGAT CACTTTATTA
2351 TAGATGCTCT ACATAAGTTT GATAAGCAGC AGACTATCGA ACAGGCATTC
2401 ACTAAAGAAC AAGATTTAGT TAAAAGATAA
  
```

The PSORT algorithm predicts inner membrane (0.16).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 85A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 85B) and for FACS analysis.

These experiments show that cp7225 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 86

The following *C.pneumoniae* protein (PID 4377248) was expressed <SEQ ID 171; cp7248>:

50
 55
 101
 151

```

1 MKFWLQCAF VGCLLLTLPC CAARRRASGE NLOQTRPIAA ANLOWESYAE
51 ALEHSKQDHK PICLFFTGSD WCMWCIMQD QILQSSEFKH FAGVHLHME
101 VDFPQKNHQP EEQRQKNQEL KAQYKVTGFP ELVFIDAEGK QLARMGFEPG
151 GGAAYVSKVK SALKLR*
  
```

A predicted signal peptide is highlighted.

55 The cp7248 nucleotide sequence <SEQ ID 172> is:

```

1 ATGAAATTTT GGTGCAAGG ATGTGCTTTT GTCGGTTGTC TGCTATTCAG
  
```

51 TTTACCTTGT TGTGCTGCAC GAAGACGTGC TTCTGGAGAA AATTGCAAC
 101 AAACCTCGTCC TATAGCAGCT GCAAATCTAC AATGGGAGAG CTATGCAGAA
 151 GCTCTTGAAC ATTCTAAACA AGATCACAAA CCTATTTGTC TTTTCTTTAC
 201 AGGATCAGAC TGGTGTATGT GGTGCATAAA AATGCAAGAC CAGATTTTGC
 251 AAAGCTCTGA GTTTAAGCAT TTTGCGGGTG TGCATCTGCA TATGGTTGAA
 301 GTTGATTTCC CCCAAAAGAA TCATCAACCT GAAGAGCAGC GCCAAAAAAA
 351 TCAAGAACTG AAAGCTCAAT ATAAAGTTAC AGGATTCCTC GAACTGGTCT
 401 TCATAGATGC AGAAGGAAAA CAGCTTGCTC GCATGGGATT TGAGCCTGGT
 451 GGTGGAGCTG CTTACGTAAG CAAGGTGAAG TCTGCTCTTA AACTACGTTA
 501 A

The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 86A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 86B) and for FACS analysis.

15 The cp7248 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7248 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 87

The following *C.pneumoniae* protein (PID 4377249) was expressed <SEQ ID 173; cp7249>:

20 1 MIPSPTPINF RDDTILETDP KPSLIMFSSK KTEIASERRK AHPTLFKVLG
 51 TIWNIVKFII SIILFLPLAL LWVLKKTCQF FILPSSIIISQ SMSKTAVAIR
 101 RMTFLSHIKQ LLSLKEISAA DRVVIQYDDL VVDSLAIKIP HALPHRWILY
 151 SQGNSGLMEN LFDRGDSSLH QLAkatGSNL LVFNYPGIMS SKGEAKRENL
 201 VKSYQACVRY LRDEETGPKA NQIIAFGYSL GTSVQAAALD REVTDGSDGT
 251 SWIVVKDRGP RSLADVANI QKPIASAIK LVGWNIDSVK PSERLRCPEI
 301 FIYNSNHDQE LISDGLFERE NCVATPFLEL PEVKTSKGTKI PIPERDLHL
 351 NPLSPNVVDR LAAVISNYLD SENRKSQQPD *

The cp7249 nucleotide sequence <SEQ ID 174> is:

30 1 ATGATCCCAT CCCCTACCCC AATAAACTTT CGTGATGATA CGATTCTAGA
 51 GACGGATCCA AAGCCGCTCT TAATCATGTT CTCTTCAAAA AAAACAGAGA
 101 TAGCTTCTGA AAGACGGAAG GCCCATCCCA CCTATTATA AGTTCTAGGA
 151 ACGATTGGGA ATATTGTGAA GTTTATTATC TCAATCATTC TGTTCTCTCC
 201 CTTAGCGTTA TTGTGGGTAC TCAAGAAAAC CTGTCAGTTT TTCATTCTCC
 251 CATCTTCTAT CATATCTCAG AGCATGTCAA AAACAGCTGT GGCAATTCCG
 301 CGAATGACCT TTCTGTCCCA TATTAAACAA CTCCTAAGCC TTAAGGAAAT
 351 CTCAGCTGCC GATCGTGTGG TTATACAATA TGACGATTG GTGGTTGATA
 401 GCTTAGCTAT AAAGATACCT CATGCTCTTC CCCACAGGTG GATTCTTTAT
 451 TCTCAAGGAA ACTCTGGATT GATGGAAAAC CTGTTTCGATC GGGGCGATTC
 501 CTCTCTACAC CAGCTAGCCA AAGCAACCGG CTCGAATCTT CTTGTGTTCA
 551 ACTATCCTGG AATTATGTCC AGCAAAGGAG AAGCGAAACG AGAAAATCTG
 601 GTTAAATCGT ATCAGGCATG CGTACGCTAC CTACGAGATG AAGAGACAGG
 651 TCCTAAAGCC AATCAAATCA TAGCTTTCGG ATACTCTTTG GGAACCTAGT
 701 TCCAAGCTGC TGCTCTAGAT CGTGAGGTCA CTGATGGCAG TGATGGAAC
 751 TCATGGATTG TTGTAAGAGA TCGGGGCCCT CGCTCTCTAG CAGATGTCCG
 801 GAATCAAAAT TGTAAGCCCA TAGCTTCCGC GATTATATAA CTCGTGGGTT
 851 GGAACATAGA CTCTGTGAAA CCTAGCGAAA GATTGCGTTG TCCCGAAAT
 901 TTCATTTACA ACTCTAATCA TGATCAAGAA CTCATTAGCG ACGGCCTCTT
 951 CGAAAGAGAA AATTGCGTAG CAACACCTTT TCTAGAGCTT CCTGAAAGTAA
 1001 AAACCTCGGG GACTAAAATT CCTATACCCG AAAGGGATCT TCTCCATCTA
 1051 AATCCTCTCA GTCCAAATGT AGTAGACAGA TTAGCAGCAG TGATCTCTAA
 1101 TTATTAGAT TCTGAAAACA GAAAGTCTCA GCAACCTGAT TAA

The PSORT algorithm predicts inner membrane (0.571).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 87A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 87B) and for FACS analysis.

These experiments show that cp7249 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 88

The following *C.pneumoniae* protein (PID 4377261) was expressed <SEQ ID 175; cp7261>:

```

1  MLPIISILFY  VILGCLSAIYI ADKKRRNVIG WFFAGAFFGF IGLVLLLLLP
51  SRRNALEKPO NDPFDNSDLF DDLKKSLAGN DEIPSSGDLQ EIVIDTEKWF
101 YLNKDRENVG PISFEELVVL LKGKTYPEEI WVKKGKMKDW QRVKDVPSLQ
151 QALKEASK*

```

The cp7261 nucleotide sequence <SEQ ID 176> is:

```

1  ATGCTCCCTA TTTCGATTTT ATTATTTTAT GTGATTCTAG GTTGTCTATC
51  TGCCTACATA GCAGATAAGA AAAAACGAAA TGTTATTGGC TGGTTTTTTG
101 CAGGAGCATT TTTTGGATT TTTGGTCTAG TTGTCCTTCT TCTTCTTCCT
151 TCTCGTCGAA ACGCTTTAGA AAAGCCACAA AACGATCCTT TTGATAACTC
201 CGATCTTTTT GATGATTTGA AAAAAAGTTT AGCAGGTAAT GACGAGATAC
251 CCTCATCGGG AGATCTTCAA GAAATCGTTA TCGATACAGA GAAGTGGTTT
301 TATTTAAATA AAGATAGAGA AAACGTAGGT CCGATATCTT TTGAGGAGTT
351 GGTCGTACTT TTAAGGGGAA AAACGTATCC AGAAGAAATT TGGGTATGGA
401 AAAAGGGAAT GAAAGATTGG CAACGAGTGA AGGATGTTCC ATCACTACAA
451 CAGGCTTTGA AAGAAGCATC AAAATAA

```

The PSORT algorithm predicts inner membrane (0.848).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 88A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 88B) and for FACS analysis.

These experiments show that cp7261 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 89

The following *C.pneumoniae* protein (PID 4377305) was expressed <SEQ ID 177; cp7305>:

```

1  MEVYSFHPAV RTSFQHRVMA ALDAWFFLGG HRLKVVSLDS CNSGWAYQEL
51  VSISTTEKVL KLLSYLLVPI VIIALLIRCL LHSNFRIDVE KERWLKIREL
101 GIDIESCKLP SSVYNQVSSF IWFEKDKSKR PRIDVDYHTL HSKDWVVFPI
151 VFQKIPKTSR FSYWFSQKET RKRDYVRNML DHVIGYLTSE GGEWLQYISK
35  TSYQSATSLD PERVLQYCLT DNQELQGEVQ RLLNEESATK SSGDKEVLLS
201 HVSDIICQW WPKFLEVIQS PAFIEELVEE VSGKLNLDLFL CLEKANTLDQ
251 ELRNSLLRAV VHHGSEGVDI KKVAGGLIY TEAIQLQIPF SRS*
301

```

The cp7305 nucleotide sequence <SEQ ID 178> is:

```

1  ATGGAAGTTT ATAGTTTTCA CCTGCGSTA AGGACTTCGT TTCAGACCGG
40  TGTAATGGCA GCACTAGATG CTGGTTTTT TCTAGGAGGG CACCGTTTAA
51  AAGTAGTTTC TCTAGATAGT TGTAACCTAG GTTGGGCGTA TCAAGAACTT
101 GTGTCTATTT CAACGACAGA AAAAGTCTTG AAACACTCTCT CTTACCTACT
151 CGTACCGATT GTCATAATAG CTCTGTTAAT TCGTTGTCTT TTACATAGCA
201 ATTTTAGGAT AGACGTAGAG AAGGAACGTT GGTTAAAAAT AAGGGAGTTA
251 GGAATTGATA TAGAAAGCTG CAAACTCCCC AGTTCTTATG TAAACCAGGT
301 TTCCTCGTTT ATTTGGTTTG AAAAAGATAA ATCCAAACGG CCACGTATTG
351 ATGTAGATTA TCATACGCTA CATAGCAAAG ACTGGGTAGT TTTCCCTATC
401

```

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

The PSORT algorithm predicts inner membrane (0.508).

15 The protein was expressed in *E. coli* and purified as a GST-fusion product (Figure 89A) and also as a double GST/his fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 89B) and for FACS analysis.

These experiments show that cp7305 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 90

20 The following *C. pneumoniae* protein (PID 4377347) was expressed <SEQ ID 179; cp7347>:

1 MKKGLGAIV FGLLTSSVA QFSKDLTKDN AYQDLNVIEH LISLKYAPLP
51 WKELLFGWDL SQQTQOARLQ LVLEEKPTTN YCQVLSNYV RSLNDYHAGI
101 TFYRTESAYI PYVLKLSLEDG HVFVVDVQTS QGDIYLGDEI LEVDGMGIRE
151 AIESLRFGRG SATDYSAAVR SLTSRSAAFG DAVPSGIAML KLRPSGLIR
201 STPVWRWYTP EHGDFSLVA PLIPEHKPQL PTQSCVLFPS GVNSQSSSSS
251 LFSSYMVPYF WEELRVQNKQ RFDSNNHIGS RNFPLPTFGP ILWEQDKGPY
301 RSYIFKAKDS QGNPHRIGFL RISSYVWTDL EGLEEDHKDS PWELFGEIID
351 HLEKETDALI IDQTHNPGGS VFYLYSLISM LTDHPLDTPK HRMIFTQDEV
401 SSSALHWQDL EDVFTDEQAV AVLGETMEGY CMDMHAVASL QNFSQSVLSS
451 WVSGDINLSK PMPLLGFAQV RPHPKHQYTK PLFMLIDEDD FSCGDLAPAI
501 LKDNGRATLI GKPTAGAGGF VFQVTFPNRS GIKGLSLTGS LAVRKDGEFI
551 ENLGVAPHID LGFTSRDLQT SRFTDYVEAV KTIVLTSLSLSE NAKKSEEQTS
601 PQETPEVIRV SYPTTTSAS*

A predicted signal peptide is highlighted.

35 The cp7347 nucleotide sequence <SEQ ID 180> is:

1 ATGAAAAAAG GGAAATTAGG AGCCATAGTT TTTGGCCTTC TATTTACAAG
51 TAGTGTGCT GGTTTTTCTA AGGATTGAC TAAAGACAAC GCTTATCAAG
101 ATTTAAATGT CATAGAGCAT TTAATATCGT TAAAAATATGC TCCTTTACCA
151 TGGAAGGAAC TATTATTGGG TTGGGATTTA TCTCAGCAAA CACAGCAAGC
201 TCGCTTGCAA CTGGTCTTAG AAGAAAAACC AACAACCAAC TACTGCCAGA
251 AGGTACTCTC TAACTACGTG AGATCATTAAC ACAGTATATCA TGCAGGGATT
301 ACGTTTTATC GTACTGAAAG TCGTATATC CTTTACGTAT TGAAGTTAAG
351 TGAAGATGGT CATGCTTTTG TAGTCGACGT ACAGACTAGC CAAGGGGATA
401 TTTACTTAGG GGATGAAATC CTTGAAGTAG ATGGAATGGG GATTCTGTAG
451 GCTATCGAAA GCCTTCGCTT TGGACGAGGG AGTGCACAG ACTATCTGC
501 TGCAGTTCGT TCCTTGACAT CGCGTCCGC CGCTTTTGA GATGCGGTTT
551 CTTCAGGAAT TGCCATGTTG AACTTCGCC GACCCAGTGG TTTGATCCGT
601 TCGACACCGG TCCGTGGCG TTATACTCCA GAGCATATCG GAGATTTTTC
651 TTTAGTTGCT CTTTGATTC CTGAACATAA ACCTCAATTA CCTACACAAA
701 GTTGTGTGCT ATTCCGTTCC GGGTAAATP CACAGTCTTC TAGTAGCTCT
751 TTATTCAGTT CCTACATGTT GCCTTATTTC TGGGAAGAAT TGCGGGTTCA
801 AAATAAGCAG CGTTTTGACA GTAATCACC AATAGGAGC CGTAATGGAT
851 TTTTACCTAC GTTTGGTCCT ATTCTTTGGG AACAAGACAA GGGGCCCTAT
901 CGTTCCTATA TCTTTAAAGC AAAAGATTCT CAGGGCAATC CCCATCGCAT
951 AGGATTTTTA AGAATTTCCT CTATGTTTG GACTGATTTA GAAGGACTTG
1001 AAGAGGATCA TAAGGATAGT CCTTGGGAGC TCTTTGGAGA GATCATCGAT

5
 10
 15

```

1051 CATTGGAAG AAGAGACTGA TGCTTTGATT ATTGATCAGA CCCATAATCC
1101 TGGAGGCAGT GTTTTCTATC TCTATTCGTT ACTATCTATG TTAACAGATC
1151 ATCCTTTAGA TACTCCTAAA CATAGAATGA TTTTCACTCA GGATGAAGTC
1201 AGCTCGGCTT TGCACGGGCA AGATCTACTA GAAGATGTCT TCACAGATGA
1251 GCAGGCAGTT GCCGTGCTAG GGGAACTAT GGAAGGATAT TGCATGGATA
1301 TGCATGCTGT AGCCTCTCTT CAAAACCTCT CTCAGAGTGT CCTTTCTTCC
1351 TGGGTTTCAG GTGATATTAA CCTTTCAAAA CCTATGCCTT TGCTAGGATT
1401 TGCACAGGTT CGACCTCATC CTAAACATCA ATATACTAAA CCTTTGTTTA
1451 TGTTGATAGA CGAGGATGAC TTCTCTTG TGAGATTAGC GCCTGCAATT
1501 TTGAAGGATA ATGGCCGCGC TACTCTCATT GGAAAGCCAA CAGCAGGAGC
1551 TGGAGGTTTT GTATTCCAAG TCACTTTCCT TAACCGTTCT GGAATTAAAG
1601 GTCTTTCTTT AACAGGATCT TTAGCTGTGA GGAAAGATGG TGAGTTTATT
1651 GAAAACCTAG GAGTGGCTCC TCATATTGAT TTAGGATTGA CCTCCAGGGA
1701 TTTGCAAACT TCCAGGTTTA CTGATTACGT TGAGGCAGTG AAACTATAG
1751 TTTTAACCTC TTTGTCTGAG AACGCTAAGA AGAGTGAAGA GCAGACTTCT
1801 CCGCAAGAGA CGCCTGAAGT TATTCGAGTC TCTTATCCCA CAACGACTTC
1851 TGCTTCGTAA
  
```

The PSORT algorithm predicts periplasmic space (0.2497).

20

The protein was expressed in *E. coli* and purified as a GST-fusion product (Figure 90A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 90B) and for FACS analysis.

These experiments show that cp7347 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 91

25

The following *C. pneumoniae* protein (PID 4377353) was expressed <SEQ ID 181; cp7353>:

30
 35

```

1 MNMPVPSAVP SANITLKEDS STVSTASGIL KTATGEVLVS CTALEGSST
51 DALISLALGQ IILATQOELL LQSTNVHQLL FLPPVEVLE IQVVDLLVQL
101 EHAETITSEP QETQTQSRSE QTL PQSSSK QSALSPRLK PEISDSKQQQ
151 ALQTPKDSAV RKHSEAPSP TQARASLSQA SSSQSRSLPP QESAPERTLL
201 EQQKASSFSP LSQFSAEKQK EALTTKSHE LYKRDQDRQ QREQHDRKHD
251 QBEDAESKSK KKKRGLGVEA VAEPEGNLD IAALIFSDQM RPPAETSKK
301 ETTFKKKLPS PMSVFSRFIP SKNPLSVGSS IHGPIQTPKV ENVFLRFMKL
351 MARILQAEA EANELYMRVK QRTDDVDLT VLI SKINNEK KDIDWSENEE
401 MKALLNRAKE IGVTDIKEY TWTEEEKRL KENVQMRKEN MEKITQMERT
451 DMQRHLQEIS QCHQARSNVL KLLKELMDTF IYNLRP*
  
```

The cp7353 nucleotide sequence <SEQ ID 182> is:

40
 45
 50
 55

```

1 ATGAATATGC CTGTTCCCTC TGCAGTTCCT TCTGCAATA TAACTCTAAA
51 AGAAGACAGC TCAACAGTTT CCACAGCCTC TGGAATATTA AAGACTGCAA
101 CAGGTGAAGT CTTAGTCTCT TGTACAGCGC TAGAAGGAAG CTCTTCTACA
151 GATGCTTTAA TTAGCTTAGC TTAGGACAA ATCATTCTTG CGACCCAACA
201 AGAACTGCTC TTACAAAGCA CAAATGTTCA TCAACTCCTC TTCCTCCCTC
251 CTGAAGTTGT AGAATTAGAA ATCCAAGTTG TTGACTTGCT AGTGCAATTG
301 GAACATGCAG AGACAATCAC AAGTGAACCA CAAGAAACAC AAACGCAAG
351 TAGGAGTGAG CAGACCCTCC CTCAACAAAG CAGCAGTAAA CAATCTGCTC
401 TCTCCCCACG CTCCTTAAAA CCTGAAATTT CTGATTCTAA ACAACAGCAA
451 GCTCTTCAAA CACCAAAAGA CTCTGCTGTA AGAAAAACA CGGAAGCACC
501 GTCACCTGAG ACACAGCTC GCGCTTCCTT ATCTCAGGCA AGCTCAAGTT
551 CTCAGAGATC CTTACCTCCG CAAGAAAGTG CGCCAGAAAG AACACTATTA
601 GAACAACAAA AAGCAAGCTC CTTCTCTCCT CTATCCAGT TCTCTGCAGA
651 GAAACAAAAA GAGGCCCTGA CGACCTCAAA ATCTCATGAA CTCTATAAAG
701 AACGCGATCA AGATCGCCAA CAAAGAGAGC AGCAGACAG AAAGCAGAT
751 CAGGAAGAAG ACCTGAATC TAAAAAGAAA AAGAAGAAAC GTGGTCTCGG
801 TGTAAGAGCA GTCGCTGAGG AACC CGGAGA AAATCTAGAT ATTGCCGCTT
851 TAACTTCTCT AGATCAAAATG CGACCTCCTG CTGAAGAAAC TTCTAAAAAA
901 GAAACGACAT TCAAAAAGAA GCTACCTTCT CCAATGTCTG TGTTTAGCAG
951 ATTCATCCCT AGTAAGAATC CGTTATCTGT AGGCTCTTCA ATACACGGGC
1001 CTATACAAAC TCCAAAAGTA GAAAATGTGT TCTTAAGGTT CATGAAGCTC
  
```

5
1051 ATGGCAAGAA TCTTAGGCCA AGCCGAAGCC GAAGCTAATG AACTCTACAT
1101 GCGAGTCAAA CAACGTACCG ATGATGTAGA CACACTCACA GTCCTTATCT
1151 CTAAGATCAA TAATGAAAAG AAAGACATTG ATTGGAGTGA AAATGAAGAG
1201 ATGAAAGCTC TTTTAAATCG AGCTAAAGAG ATTGGAGTCA CTATAGACAA
1251 AGAAAAATAT ACTTGGACAG AAGAGGAAAA AAGACTTCTA AAAGAGAATG
1301 TCCAAATGCG CAAAGAGAAAT ATGGAGAAAA TCACTCAAAT GGAAAGGACG
1351 GACATGCAAA GGCACCTCCA AGAGATTTCT CAATGTCATC AAGCGCGCTC
1401 TAATGTATTG AAGTTATTGA AAGAATTAT GGACACCTTC ATTTACAACC
1451 TACGCCCTA A

10 The PSORT algorithm predicts cytoplasm (0.1308).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 91A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 91B) and for FACS analysis.

15 These experiments show that cp7353 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 92

The following *C.pneumoniae* protein (PID 4377408) was expressed <SEQ ID 183; cp7408>:

20
1 MLKIQQKRCM VSVVITVGAI VGFFNSADAA PKKKKIPIQI LYSFTKVSSY
51 LKNEDASTIF CVDVDRGLLQ HRYLGSPGWQ ETRRRQLFKS LENQSYGNER
101 LGEETLAIDI FRNKECLESE IPEQMEAILA NSSALVLGIS SFGITGIPAT
151 LHSLLRQNLN FQKRISASES FLLKIDSAPS DASVFYKGV LFRGETAIVDA
201 LSQLEAQLDL SPKKIIFLGE DPEVVQAVGS ACIGWGMNFI GLVYYPQES
251 LFSYVHPYST ATELQEAQGL QVISDEVAQL TLNALPKMN*

The cp7408 nucleotide sequence <SEQ ID 184> is:

25
1 ATGTTGAAAA TCCAGAAAAA AAGAAATGTGT GTCAGCGTAG TCATCACGGT
51 AGGCGCCATA GTGGGGTTTTC TCAATCTGTC AGACGCAGCA CCAAGAAAAA
101 AGAAGATCCC TATACAGATT CTCCTACTCCT TTAATAAGT CTCTTCCTAT
151 TTAAAAAACG AAGACGCAAG TACTATATTT TGCGTCGATG TGGATCGTGG
201 ACTTCTCCAG CATCGGTATT TAGGTAGTCC AGGATGGCAG GAAACCAGAC
251 GTCGGCAGTT ATTTAAATCC TTAGAAAATC AATCATACGG CAACGAACGT
301 TTAGGAGAAG AAACCTCTGC TATTGATATT TTCAGGAACA AAGAGTGCTT
351 GGAGAGCGAG ATCCAGAGC AGATGGAAGC TATCCTTGCA AATTCCTCGG
401 CCTTGGTCTT AGGCATCTCT TCTTTTGGA TCACAGGAAT TCCTGCGACT
451 TTGCATAGTT TGCTTCGACA GAATCTATCT TTCCAAAAAC GCTCTATAGC
35 501 ATCGGAGAGC TTCCTTTTAA AGATCGATAG TGCCCCCTCA GATGCCCTCG
551 TTTTATATAA AGCGGTGCTT TTCCGCGGAG AGACTGCGAT CGTGGATGCG
601 TTAAGCCAAT TATTGCCCCA GCTCGATCTT TCTCCTAAAA AAATTATCTT
651 TCTAGGAGAA GACCGTGAGG TCGTTCAAGC TGTGGGTCT GCTTGTATAG
701 GTTGGGCGAT GAACTTTTAA GGCCTGGTAT ACTATCCTGC TCAAGAAAGC
40 751 CTTTTTCTT ATGTTTCATCC TTAATCTACA GCAACGAGC TCCAAGAAGC
801 ACAGGGTTTA CAAGTAATTT CAGATGAAGT CGCACAGCTT ACTTTAAACG
851 CTCTTCCGAA AATGAATTAA

The PSORT algorithm predicts inner membrane (0.123).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 92A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 92B) and for FACS analysis.

These experiments show that cp7408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 93

The following *C.pneumoniae* protein (PID 4376424) was expressed <SEQ ID 185; cp6424>:

```

1 MMHNIVVLSE EPGRSAFLGR TAFFPNKYPI AQGGVGIPST IGNLFTIWCY
51 FYFYRAATPQ SDHPDGCGFI LLERLKLGA GFFYCDLRES NTTGFTLFFE
101 GSNKGVKLNH LFIRDE*

```

The cp6424 nucleotide sequence <SEQ ID 186> is:

```

1 ATGATGCACA ATATTGTTGT TCTTAGTGAG GAACCTGGAC GAAGCGCTTT
51 TCTTGGTAGG ACGGCATTTT TCCCTAATAA GTATCCAATA GCTCAGGGTG
101 GTGTTGGAAT ACCATCTACA ATAGGCAATC TCTTACTAT ATGGTACTGT
151 TTCTATTTT ATAGAGCTGC AACTCCACAA TCTGATCATC CTGACGGATG
201 TGGCTTTATT CTAAGAGAAA GGCTTAAGGA GCTCGGTGCA GGGTTCTTTT
251 ATTGTGATCT TCGTGAGTCC AATACCACTG GCTTACTCT TTTTGTGAA
301 GGCTCCAATA AAGGTGTGTT AAAGAATCAC TTGTTTATTA GAGATGAGTA
351 A

```

The PSORT algorithm predicts cytoplasm (0.2502).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 93A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 93B) and for FACS analyses (Figure 93C; GST-fusion).

These experiments show that cp6424 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 94

The following *C.pneumoniae* protein (PID 4376449) was expressed <SEQ ID 187; cp6449>:

```

1 VASETYPYQI LHAQREVRDA YFNQADCHPA RANQILEAKK ICLLDVYHTN
51 HYSVFTFCVD NYPNLRFTFV SSKNEMNGL SNPLDNVLVE AMVRRTHARN
101 LLAACKIRNI BVPRVGLDL RSGILISKLE LKQPQFQSLT EDFVNHSTNQ
151 BEARVHQKHV LLISLILLCK QAVLESFQEK KRSS*

```

The cp6449 nucleotide sequence <SEQ ID 188> is:

```

1 GTGGCGTCTG AAACGTATCC TTCTCAGATA TTGCACGCTC AGAGGGAAGT
51 ACGTGATGCC TATTTTAATC AAGCGGATTG CCATCCTGCT CGGGCTAATC
101 AGATTCTCGA GGCTAAGAAA ATCTGTTTAT TAGATGTTTA TCATACTAAT
151 CATTATTCCG TATTACTTTT TTGTGTAGAT AATTATCCGA ATCTCCGCTT
201 TACATTTGTA TCTTCAAAAA ACAATGAGAT GAATGGCTTA TCTAATCCTC
251 TAGATAATGT TCTTGTAGAG GCTATGGTAC GTAGAACACA TGCAAGAAAC
301 CTACTTGACG CGTGTAATAA TCGAAATATT GAGGTTCCTA GGGTGTGTTG
351 GCTTGACCTA AGATCTGGGA TACTCATTTT GAAACTAGAA TTGAAGCAAC
401 CTCAGTTCCA AAGTTTAACA GAAGACTTCG TAAATCATTC CACAAATCAG
451 GAAGAAGCTC GCGTCCATCA AAAGCATGTG TTGCTAATTT CTTTAATTTT
501 ACTTTGCAAG CAGGCCGTTT TGGAATCATT CCAGGAAAAA AAGCGATCCT
551 CTTAA

```

The PSORT algorithm predicts inner membrane (0.2084).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 94A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 94B) and for FACS analyses (Figure 94C; GST-fusion).

These experiments show that cp6449 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 95

The following *C.pneumoniae* protein (PID 4376495) was expressed <SEQ ID 189; cp6495>:

MRELNAFELTQPEEYRNRWVLMPCCLKRFRCTQHAKVWSYRCVHEASLYEKNCFLLTLYDDKHL PQYGSVLVHLQLFLKR
LRKMISPHKIRYFECGAYGTKLQRPYHLLLS

- 5 The cp6495 nucleotide sequence <SEQ ID 190> is:

TTGCGAGAATTAAATGCTTTTGAATTAACCAACCTGAAGAGTATCGAAACCGTTGGGTTTTGATGCCTTGTCTTAAGTGT
CGTTTTTTGTAGAACGCAACATGCAAAAGTCTGGTCTTATCGTTGTGTCCATGAAGCTTCTTTGTATGAGAAAAATTGTTTT
CTTACTTTGACTTATGATGATAAGCATTACCTCAGTATGGTTGGTAAAGCTGCATTACAGCTGTTTCTTAAGAGA
TTAAGAAAGATGATTTCTCCTCATAAAATTCGTATTTTGAATGTGGTGCATGGAACCAATTACAAAGACCTCATATAT
CATCTACTTTTATCATGA

The PSORT algorithm predicts cytoplasmic (0.280).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 95A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 95B) and for FACS analysis (Figure 95C).

- 15 These experiments show that cp6495 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 96

The following *C.pneumoniae* protein (PID 4376506) was expressed <SEQ ID 191; cp6506>:

1 MRRFLFLILS SLPLVAFSAD NPTILEEKQS PLSRVSIIFA LPGVTPVSFD
51 GNCPIPWFSH SKKTLEGQRI YYSGDSFGKY FVVSALWPNK VSSAVVACNM
101 ILKHRVDLIL IIGSCYSRSQ DSRFGSVLVS KGYINYDADV RPFERFEIP
151 DIKKS VFATS EVHREAILRG GEEFI STHKQ EIEELLKTHG YLKSTTKTEH
201 TLMEGLVATG ESFAMSRNYF LSLQKLYPEI HGFDSVSGAV SQVCYEYSIP
251 CLGVNILLPH PLESRSNEDW KHLQSEASKI YMDTLKSVL KELCSSH*

- 25 The cp6506 nucleotide sequence <SEQ ID 192> is:

1 ATGCGTCGTT TTCTGTTTCT TATTCTTAGC TCTCTTCCTT TGGTCGCATT
51 CTCTGCTGAT AATTTCACTA TTCTAGAAGA AAAACAGAGT CCTTTAAGTC
101 GTGTAAGTAT TATTTTGGCT TTACCTGGGG TTACTCCCGT TTCTTTTGAT
151 GGTAAATTGTC CTATTCCTTG GTTTTCTCAT AGTAAAAAGA CTCTAGAGGG
201 ACAGAGAATT TATTA CTCTG CCGACTCCTT TGGGAAATAC TTTGTAGTTT
251 CTGCTCTTTG GCCTAATAAA GTTCTTTCAG CTGTTGTGGC TTGTAATATG
301 ATTCTTAAAC ATCGAGTGGA TCTTATTCTA ATTATAGGCT CGTGTACTCT
351 TAGGTCTCAA GATAGCCGTT TTGGCAGCGT CTTAGTTTCT AAAGGCTACA
401 TTAATTATGA TGCAGATGTG AGGCCTTTCT TTGAAAGATT TGAGATTCCA
451 GACATTAAAA AGAGTGTTTT TGCAACCACT GAGGTTTCAT GGGAGGCAAT
501 TCTTCGTGGA GGCGAAGAGT TTATTTCTAC CCATAAACAA GAAATCGAAG
551 AGCTTTTGAA GACTCATGGG TATTTGAAAT CAACAACCAA AACGGAGCAC
601 ACCTTAATGG AAGGTTTGGT TGCTACAGGC GAGTCTTTCG CGATGTCGCG
651 AAATAATTTT CTTTCCTTAC AAAAATTGTA TCCAGAGATT CATGGTTTGT
701 ATAGTGTGAG CGGCGCTGTT TCTCAGGTAT GCTATGAATA TAGCATTCCT
751 TGTTTAGGTG TGAATATCCT TCTCCCTCAT CCTTTAGAA CACGGAGTAA
801 CGAGGATTGG AAGCATCTTC AAAGTGAGGC AAGTAAATTT TATATGGATA
851 CCTTGCTCAA GAGTGATTAT AAAGA ACTCT GTTCTTCTCA TTAA

The PSORT algorithm predicts periplasmic space (0.571).

- 45 The protein was expressed in *E.coli* and purified as his-tag (Figure 96A) and GST-fusion (Figure 96B) products. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 96C) and for FACS analysis (Figure 96D).

These experiments show that cp6506 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 97

The following *C.pneumoniae* protein (PID 4376882) was expressed <SEQ ID 193; cp6882>:

```

5      1  MSLLNLPSSQ DSASEDSTSQ SQIFDPINR ELVSTPEEKV RQRLLSFLMH
      51  KLNYPKKLII IEKELKTLFP LLMRKGTLP KRRPDILIT PPTYTDAQGN
     101  THNLGDPKPL LLIECKALAV NQNALQQLS YNYSIGATCI AMAGKHSQVS
     151  ALFNPKTQTL DFYPLPEYS QLLNYFISLN L*
  
```

The cp6882 nucleotide sequence <SEQ ID 194> is:

```

10      1  ATGTCCTTAT TGAACCTTCC CTCAGCCAG GATTCTGCAT CTGAGGACTC
      51  CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
     101  CTACTCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
     151  AAGCTGAAC TACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAAC
     201  TCTTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
     251  CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
     301  ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
     351  CTTAGCCGTA AACCAAAATG CACTCAACA ACTCCTTAGC TATAACTACT
     401  CTATCGGAG CACCTGCATT GCTATGGCAG GGAACACTC TCAAGTGTC
     451  GCTCTCTTCA ATCCAAAAAC ACAAACTCTT GATTTTATC CTGGCCTCCC
     501  AGAGTATTC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG
  
```

The PSORT algorithm predicts cytoplasm (0.362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 97A). The protein was used to immunise mice, whose sera were used in a Western blot (Figure 97B) and for FACS analysis (Figure 97C).

25 These experiments show that cp6882 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 98

The following *C.pneumoniae* protein (PID 4376979) was expressed <SEQ ID 195; cp6979>:

```

30      1  MSVNPSGNSK NDLWITGAHD QHPDVKESGV TSANLGSHRV TASGGRQGLL
      51  ARIKEAVTGF FSRMSFFRSQ APRGSQQPSA PSADTVRSPL PGGDARATEG
     101  AGRNLIKKG YQPMKVITPQ VPGGGAQRSS GSTTLKPTRP APPPPKTGGT
     151  NAKRPATHGK GPAPQPPKTG GTNAKRAATH GKGPAPQPPK GILKQPGQSG
     201  TSGKKRVSW DED*
  
```

The cp6979 nucleotide sequence <SEQ ID 196> is:

```

35      1  ATGTCTGTTA ATCCATCAGG AAATCCAAG AACGATCTCT GGATTACGGG
      51  AGCTCATGAT CAGCATCCCC ATGTAAAGA ATCCGGGGTT ACAAGTGCTA
     101  ACCTAGGAAG TCATAGAGTG ACTGCCTCAG GAGGACGCCA AGGGTTATTA
     151  GCACGAATCA AAGAAGCAGT AACCGGGTTT TTTAGTCGGA TGAGCTTCTT
     201  CAGATCGGGA GCTCCAAGAG GTAGCCAACA ACCCTCTGCT CCATCTGCAG
     251  ATACTGTACG TAGCCCGTTG CCGGGAGGGG ATGCTCGCGC TACCGAGGGA
     301  GCTGGTAGGA ACTTAATTAA AAAAGGGTAC CAACCAGGGA TGAAAGTCAC
     351  TATCCCACAG GTTCTCTGGG GAGGGGCCCA ACGTTCAATCA GGTAGCACGA
     401  CACTAAAGCC TACCGTCCG GCACCCCCAC CTCCTAAAAC GGGTGGAAC
     451  AATGCAAAAC GTCCGGCAAC GCACGGGAAG GGTCCAGCAC CCCAGCCTCC
     501  TAAACAGGT GGGACCAATG CTAAGCGCGC AGCAACGCAT GGGAAAGGTC
     551  CAGCACCTCA ACCTCCTAAG GGCATTTTGA AACAGCCTGG GCAGTCTGGG
     601  ACTTCAGGAA AGAAGCGTGT CAGCTGGTCT GACGAAGATT AA
  
```

The PSORT algorithm predicts cytoplasm (0.360).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 98A). The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 98B) and for FACS analysis (Figure 98C).

These experiments show that cp6979 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 99

The following *C.pneumoniae* protein (PID 4377028) was expressed <SEQ ID 197; cp7028>:

```

1  MLLGFLCDP CASWQCAAVA NCYDSVFMSR PEHKPNIPYI TKATRRGLRM
51  KTLAYLASLK DARQLAYDFL KDPGSLARLA KALIAPKEAL QEGNLFYFGC
101 SNIEDILEEM RRPHRILLG FSYCQKPKAC PEGRFNDACR YDP SHPTCAS
151 CSIGTMMRLN ARRYTTVILP TFDIAKHLH TLKKRYPGYQ ILFAVTACEL
201 SLKMFGDYAS VMNLKGVGIR LTGRICNTFK AFKLAERGVK PGVTILEEDG
251 FEVLARILTE YSSAPFPRDF CEIH*

```

The cp7028 nucleotide sequence <SEQ ID 198> is:

```

1  ATGCTTCTAG GGTTTTGTG TGA CTGCCCC TGTGCTTCGT GGCAGTGTGC
51  GGCCGTTGCT AATTGTTATG ATTCCGTATT TATGCTCTAGA CCAGAGCACAC
101 AACCTAATAT TCCTTATATT ACTAAAGCTA CAAGACGGGG TCTGCGTATG
151 AAGACGCTTG CTTATCTGGC CTCTTTAAAA GATGCTAGAC AGCTTGCCTA
201 TGATTTTCTG AAAGATCCTG GTTCTTTAGC TCGGTTAGCT AAGGCTTTGA
251 TAGCTCCTAA GGAGGCCTTA CAGGAGGGCA ACCTATTTTT TTATGGCTGT
301 AGTAATATTG AGGATATTTT AGAGGAGATG CGTCGTCCTC ATAGAATCCT
351 TTTGTTAGGA TTTTCTTATT GTCAAAAGCC TAAGGCATGT CCTGAAGGCC
401 GTTTCATGA TGCTTGTCGG TATGATCCTT CACATCCTAC ATGTGCCTCA
451 TGTTCTATAG GGACCATGAT GCGGCTGAAT GCTCGTAGAT AACTACTGT
501 GATCATCCCT ACATTATAG ATATCGCAA ACATTACAC ACTTTAAAAA
551 AGCGCTACCC TGGATATCAA ATTCTCTTTG CAGTTACTGC TTGTGAACCT
601 TCCTTAAAAA TGTTTGGAGA TTATGCCTCC GTAATGAAC TAAAGGGTGT
651 GGGCATCAGA CTCACAGGAC GTATTGCAA TACATTTAAG GCATTAAAT
701 TAGCTGAGCG AGGAGTCAAA CCAGGAGTCA CTATCTAGA AGAAGATGGC
751 TTTGAGGTAT TAGCAAGGAT TCTTACAGAA TACAGTAGCG CTCCTTTCCC
801 TAGAGACTTT TGTGAGATCC ATTAG

```

The PSORT algorithm predicts cytoplasm (0.1453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 99A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 99B) and for FACS analysis (Figure 99C).

These experiments show that cp7028 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 100

The following *C.pneumoniae* protein (PID 4377355) was expressed <SEQ ID 199; cp7355>:

```

1  MKKVVTLSII FFATYCASEL SAVTVVAVPL SEAPGKIQVR PVVGLQFQEE
51  QGVPYSPFY PYDGYGYYPE TYGYTKNTGQ ESRECYTRFE DGTIFYECD*

```

The cp7355 nucleotide sequence <SEQ ID 200> is:

```

1  ATGAAGAAAG TCGTAACACT ATCCATTATA TTTTTCGCAA CGTATTGTGC
51  ATCAGAGCTT AGTGCTGTAA CTGTAGTGGC TGTGCCCTTA TCAGAGGCTC
101 CAGGGAAGAT TCAAGTTCGT CCCGTCGTTG GTCTGCAATT TCAAGAAGAA
151 CAGGGTTCTG TGCCCTATAG TTTTATTAT CCTTATGACT ATGGGTATTA
201 CTATCCAGAG ACTTATGGCT ATACTAAAAA TACAGGTCAA GAAAGTCGCG

```

251 AATGTTATAC CCGATTGAA GATGGCACAA TTTTATGA ATGCGATTAG

The PSORT algorithm predicts inner membrane (0.143).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 100A) and a his-tag product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 100B) and for FACS analysis (Figure 100C).

These experiments show that cp7355 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 101

The following *C.pneumoniae* protein (PID 4377380) was expressed <SEQ ID 201; cp7380>:

```

10      1  VHYCERTLDP KYILKIALKL RQSLSLFFQN SQSLQRAYST PYSYYRIILQ
      51  KENKEKQALA RHKCISILEF FKNLLFVHLL SLKSNQREGC STDMAVSTP
      101 FFNRLWYRL LSSRFSWLKS YCPRFFLDYL EAFGLLSDFL DHQAVIKFFE
      151 LETHFSYYPV SGFVAPHNL SLLQDRYFPI ASVMRTLDDK NFSLTPDLIH
      201 DLLGHVFWLL HPSFSEFFIN MGRLFKIVIE KVQALPSKKQ RIQTLQSNLI
      251 AIVRCFWFTV ESLGIENHEG RKAYGAVLIS SPQELGHAFI DNVRLPLEL
      301 DQIILPFNT STPQETLPSI RHFDELVELT SKLEWMLDQG LLESTPLYNQ
      351 EKYLSGFEVL CQ*
```

The cp7380 nucleotide sequence <SEQ ID 202> is:

```

20      1  GTGCACTACT GCGAGAGAAC CCTGGACCCA AAGTATATTC TGAAGATTGC
      51  TCTAAAGCTG AGACAATCAC TTTCCCTGTT CTTCCAGAAC AGCCAATCAC
      101 TCCAACGTGC ATACTCGACC CCATATTCCT ACTACCGAAT CATTCTACAA
      151 AAGGAAAATA AAGAGAAGCA AGCTTAGCT CGACACAAAT GCATTCTAT
      201 TTTAGAATTT TTCAAAACT TACTCTTGT TCATCTCTG TCATTATCAA
      251 AGAATCAAAG GGAAGGTTGC TCCACTGATA TGGCTGTTGT AAGCACTCCC
      301 TTTTAAATC GGAATTTATG GTATCGACTC CTTTCCTCAC GGTTTCTCT
      351 ATGGAAGAGC TATGTGCCAA GATTTTCTT TGATTACTTA GAAGCTTTCCG
      401 GTCTCCTTTC TGATTCTCTA GACCATCAAG CAGTCATTAA ATTCTTCGAA
      451 TTAGAAACAC ATTTTCTCTA TTATCCCGTT TCAGGATTG TAGCTCCCCA
      501 TCAATACTTG TCTCTGTTGC AGGACCGTTA CTTTCCATT GCCTCTGTAA
      551 TGCGAATCTT CGATAAAGAT AATTCTCTCT TAATCCTGA TCTCATCCAT
      601 GACCTTTTAG GGCACGTGCC TTGGCTTCTA CATCCCTCAT TTTCTGAATT
      651 TTTTATAAAC ATGGGAAGAC TCTTCACTAA AGTCATAGAA AAAGTACAAG
      701 CTCTTCTTAG TAAAAACAA CGCATACAAA CCTACAAAG CAATCTGATC
      751 GCTATTGTAC GCTGCTTTTG GTTACTGTT GAAAGCGGAC TTATTGAAAA
      801 CCATGAAGGA AGAAAAGCAT ATGGAGCCGT TCTTATCAGT TCTCCTCAGG
      851 AACTTGGACA CGCTTTCATT GATAACGTAC GTGTTCTCCC TTAGAATTG
      901 GATCAGATTA TTCGTCTTCC CTCAATACA TCAACTCCAC AAGAGACTTT
      951 ATTTTCAATA AGACATTTG ATGAACGGT AGAACTCACT TCAAAATTAG
      1001 AATGGATGCT CGACCAAGGT CTGTTAGAA CAATTCCCTT TTACAATCAA
      1051 GAGAAATATC TTTCTGTTT TGAGGTACTT TGCCAATGA
```

The PSORT algorithm predicts inner membrane (0.1362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 101A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 101B) and for FACS analysis (Figure 101C).

These experiments show that cp7380 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 102

The following *C.pneumoniae* protein (PID 4376904) was expressed <SEQ ID 203; cp6904>:

1 MMNYEDAKLR QQAVAILYQI GAIKFGKHIL ASGEETPLYV DMRLVISSPE
 51 VLQTVATLIW RLRPSFNSSL LCGVPYTALT LATSISLKYN IPMVLRRKEL
 101 QNVDPDAIK VEGLFPPGQT CLVINDMVSS GKSI IETAVA LEENGLVVRE
 151 ALVFLDRRKE ACQPLGPQGI KVSSVFTVPT LIKALIAYGK LSSGDLTLAN
 201 KISEILEIES *

The cp6904 nucleotide sequence <SEQ ID 204> is:

1 ATGATGAAC ACGAAGATGC AAAATTACGC GGTCAAGCTG TAGCAATTCT
 51 ATACCAAATC GGAGCTATAA AGTTCGGAAA ACATATTCTC GCTAGCGGAG
 101 AAGAACTCC TCTGTATGTA GATATGCGTC TTGTGATCTC CTCTCCAGAA
 151 GTTCTCCAGA CAGTGGCAAC TCTTATTTGG CGCTCCGCC CCTCATTCAA
 201 TAGTAGCTTA CTCTGCGGAG TCCCTTATAC TGCTCTAACC CTAGCAACCT
 251 CGATCTCTTT AAAATATAAC ATCCCTATGG TATTGCGAAG GAAGGAATTA
 301 CAGAATGTAG ACCCTCGGA CGCTATTAAA GTAGAAGGGT TATTTACTCC
 351 AGGACAACT TGTTTAGTCA TCAATGATAT GGTTCCTCA GGAAATCTA
 401 TAATAGAGAC AGCAGTCGCA CTGGAAGAAA ATGGTCTGGT AGTTCGTGAA
 451 GCATTGGTAT TCTTAGATCG TAGAAAAGAA GCGTGTCAAC CACTTGGTCC
 501 ACAGGGAATA AAAGTCAGTT CGGTATTTAC TGTACCCACT CTGATAAAG
 551 CTTTGATCGC TTATGGGAAG CTAAGCAGTG GTGATCTAAC CCTGGCAAAC
 601 AAAATTTCCG AAATCTAGA AATTGAATCT TAA

20 The PSORT algorithm predicts cytoplasm (0.0358).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 102A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 102B) and for FACS analysis.

The cp6904 protein was also identified in the 2D-PAGE experiment.

25 These experiments show that cp6904 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 103

The following *C.pneumoniae* protein (PID 4376964) was expressed <SEQ ID 205; cp6964>:

30 1 MKKLIALIGI FLVPIKNTN KEHDAHATVL KAARAKYNLF FVQDVFPVHE
 51 VIEPISPDCL VHYEGWV*

The cp6964 nucleotide sequence <SEQ ID 206> is:

1 ATGAAAAAAT TGATTGCTTT GATAGGGATA TTTCTGTGTC CAATAAAAGG
 51 AAATACCAAT AAGGAACACG ACGCTCACGC GACTGTTTTA AAAGCGGCCA
 101 GAGCAAAGTA TAATTTGTTC TTTGTTTCAGG ATGTTTCCC TGTACACGAA
 151 GTTATCGAGC CTATTTCTCC CGATTGCCTG GTACATTATG AAGGGTGGGT
 201 TTGA

The PSORT algorithm predicts inner membrane (0.091).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 103A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 103B) and for FACS analysis (Figure 103C).

40 These experiments show that cp6964 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 104

The following *C.pneumoniae* protein (PID 4377387) was expressed <SEQ ID 207; cp7387>:

1 LNFPAKIDHNH LYLTCGLDGL VACPILSTDC LPNYSEKASH EVLVYSKFRG
 51 ISGEPSRLAT SGNDTYYSIV SLPGLRYEV TSPSGRHDFN IDMHVAPKIG
 101 AVLSHGTRTA KEIPGSSKDY AFFSLTARE LMISEKLAMT FQVSEVIQNC
 151 YSQCTKVTKT NLKEQYRHLN HNTGFELSVK SAF*

5 The cp7387 nucleotide sequence <SEQ ID 208> is:

1 TTGAATTTTG CAAAGATTGA TCACAATCAT CTCTACCTTA CATGTTTGGG
 51 AGATCTTGGT GTAGCTTGTC CTATACTTTC TACAGATTGT CTACCTAATT
 101 ATAGCGAGAA AGCATCTCAT GAGGTCTTGG TTTATAGTAA ATTTAGATGC
 151 ATTTCTGGAG AGCCATCTCG ACTTGCAACT TCAGGAAATG ACACATATTA
 10 TTTCTATAGTA AGTTTACCTA TAGGACTCCG TTACGAAGTG ACTTCACCAT
 201 CAGGACGTCA TGATTTCAT ATTGATATGC ATGTAGCTCC AAAGATAGGT
 251 GCAGTACTCT CTCATGGAAC ACGAGAGGCT AAAGAGATCC CAGGATCTTC
 301 AAAAGACTAT GCATTTTTTA GCTTGACTGC TAGAGAAAGT TTAATGATTT
 351 CTGAAAAGCT TCGCATGACT TTCCAAGTTA GCGAAGTTAT TCAGATTGT
 401 TATTACAAT GTACTAAAGT AACGAAAAC AATTAAAG AACAGTATAG
 15 451 GCACTTATCC CACAATACAG GGTGTGAGTT AAGCGTCAAG TCTGCATTCT
 501 AA

The PSORT algorithm predicts inner membrane (0.043).

The protein was expressed in *E.coli* and purified as a his-tagged-fusion product (Figure 104A) and
 20 also as a GST-fusion (Figure 104B). The recombinant proteins were used to immunise mice, whose
 sera were used in a Western blot and for FACS analysis (Figure 104C; his-tagged).

These experiments show that cp7387 is a surface-exposed and immunoaccessible protein, and that it
 is a useful immunogen. These properties are not evident from the sequence alone.

Example 105

25 The following *C.pneumoniae* protein (PID 4376281) was expressed <SEQ ID 209; cp6281>:

1 MFLQFFHPIV FSDQSLSFLP YLGKSSGIE KCSNIVEHYL HLGDDTSVII
 51 TGVSGATFLS VDHALPISKS EKIKILSYI LILPLILALF IKIVLRIILF
 101 FKYRGLILDV KKEDLKKTLT PDQENLSLPL PSPTTLKKIH ALHILVRSKG
 151 TYNELIQEGF SFTKITDLGQ APSPKQDIGF SYNLLPNFY FHSLSVSPNI
 30 201 SGEERALNYH KEQQEEMAVK LKTMQACSFV FRSLHLPSMQ TKDKKAGFGL
 251 LTFPPWKIYP L*

The cp6281 nucleotide sequence <SEQ ID 210> is:

1 ATGTTTCTTC AGTTTTCCTA TCCTATAGTC TTCTCGGATC AGTCCTTATC
 35 51 TTTTCTTCCT TACCTAGGAA AAAGCTCTGG CATTATTGAA AAATGTTCCA
 101 ATATCGTTGA AACTATTGTA CATTGTTGGG GAGACACTTC TGTATGATC
 151 ACAGGAGTTT CTGGAGCTAC CTTTCTATCT GTTGATCATG CCTCCCAAT
 201 CTCGAAATCT GAAAAAATAA TAAAAATTCT CTCCTATATT TTAATCTTTC
 251 CTCTGATTCT AGCTCTCTTT ATTAAGATCG TTTTACGCAT TATCTTATTC
 301 TTCAAGTATC GTGGTCTAAT CCTAGATGTT AAGAAGGAGG ATTTGAAAAA
 351 AACACTTACA CCTGACCAAG AAAACCTCAG TCTTCCTTTA CCATCTCCTA
 401 CAACATTAAA GAAAATTCAT GCGCTACACA TTTTAGTGCG TTCTGGAAAA
 451 ACCTATAACG AGCTTATACA AGAAGGGTTT TCTTTCCTTA AATCACAGA
 501 TCTTGGTCAA GCTCCTTCAC CAAAGCAAGA TATTGGCTTC TCTTATAATT
 551 CCTTCTCTCC TAACCTCTAT TTTCATTCCCT TGGTATCTGT TCCAAATATT
 45 601 TCAGGCGAGG AACGGGCTCT TAATTATCAT AAAGAACAAC AAGAGGAAAT
 651 GGCTGTAAA TAAAAACAA TGCAAGCGTG TTCTTTTGTG TTCGATCCC
 701 TGCATTTACC TTCAATGCAA ACGAAGGACA AAAAGGCTGG ATTTGGACTA
 751 CTGACGTTT TCCCTTGGA AATCTACCCC CTATAA

The PSORT algorithm predicts inner membrane (0.5373).

50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 105A). The
 recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure
 105B) and for FACS analysis.

These experiments show that cp6281 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 106 and Example 107

- 5 The following *C.pneumoniae* protein (PID 4376306) was expressed <SEQ ID 211; cp6306>:

```

1  MGNHETYIHP  GVLPSHAQD  VSRSTVYPSR  SFIMRRMLMG  WNFNRVPSKS
51  SEQLMDGHRI  PLIFFGKHPH  TISILNVNRF  SWLSIFYNGE  RGF*

```

The cp6306 nucleotide sequence <SEQ ID 212> is:

```

10      1  ATGGGAAACC  ATGAGACCTA  TATACATCCA  GGAGTGCTCC  CGAGTAGTCA
      51  TGCTCAGGAT  GTTAGCAGAT  CTACAGTTTA  CCCCAGTCGA  AGTTTATCA
      101  TGAGACGTAT  GCTCATGGGC  TGGAAATTCA  ATCGTGTTCC  CTCGAAGAGC
      151  TCCGAGCAGT  TAATGGATGG  TCATCGCATA  CCTCTTATAT  TTTTGGGAA
      201  GCATCATCCT  ACTATATCTA  TTTTAAATGT  CAATAGATTT  TCTTGGCTCT
      251  CCATTTTSTA  CAATGGAGAA  AGGGGGTTTT  GA

```

- 15 The PSORT algorithm predicts cytoplasm (0.167).

The following *C.pneumoniae* protein (PID 4376434) was also expressed <SEQ ID 213; cp6434>:

```

1  MSESINRSIH  LEASTPFFIK  LTNLCEsRLV  KITSLVISLL  ALVGAGVTLV
51  VLFVAGILPL  LPVLILEIIL  ITVLVLLFCL  VLEPYLIEKP  SKIKELPKVD
101  ELSVVETDST  L*

```

- 20 The cp6434 nucleotide sequence <SEQ ID 214> is:

```

      1  ATGTCTGAAA  GTATTAACAG  AAGCATTCAT  TTAGAAGCCT  CTACACCATT
      51  TTTTATAAAA  TTAACGAATC  TCTGTGAAAG  TAGATTAGTT  AAGATCACTT
      101  CTCTGTGTAT  TTCTCTATTA  GCTTTAGTGG  GTGCGGGAGT  CACTCTTGTC
      151  GTTTTATTTG  TAGCTGGGAT  CCTTCTTTTA  CTTCTGTGAC  TCATCTTAGA
      201  AATTATTTTA  ATAACCGTCC  TTGTCTTGCT  TTTTGTGTTG  GTATTGGAAC
      251  CTATTTAAT  AGAAAAACCT  AGTAAATAA  AGGAACTACC  TAAAGTAGAC
      301  GAGCTATCTG  TAGTAGAAAC  GGACAGTACT  CTTTAA

```

The PSORT algorithm predicts inner membrane (0.6859).

- 30 The proteins were expressed in *E.coli* and purified as his-tag products (Figure 106A; 6306 = lanes 2-4; 6434 = lanes 8-10). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 106B & 107) and for FACS analysis.

These experiments show that cp6306 & cp6434 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from the sequences alone.

Example 108

- 35 The following *C.pneumoniae* protein (PID 4377400) was expressed <SEQ ID 215; cp7400>:

```

1  MRVMRFFCLF  FLGFLGSFHC  VAEDKGVDLF  GVWDDNQITE  CDDSYMTEGR
51  EEVEKVVD

```

The cp7400 nucleotide sequence <SEQ ID 216> is:

```

40      1  GTGAGAGTTA  TGAGATTTT  TTGTCTATTT  TTTCTTGGGT  TCCTAGGATC
      51  TTTTCATTGT  GTTGCTGAAG  ACAAGGCGGT  GGATTATTT  GGAGTCTGGG
      101  ACGATAACCA  AATTACAGAG  TGTGACGATA  GTTACATGAC  AGAGGGTCGT
      151  GAAGAGGTTG  AAAAGGTAGT  GGACGCTTAG

```

The PSORT algorithm predicts periplasmic space (0.924).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 108A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 108B) and for FACS analysis.

These experiments show that cp7400 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 109

The following *C.pneumoniae* protein (PID 4376395) was expressed <SEQ ID 217; cp6395>:

```

1  MENAMSSSFV YNGPSWILKT SVAQEVFKKH GKGIQVLLST SVMFLFIGLV
51  CAFIFPQYLI VFVLTIALLM LAISLVLFLL IRSVRSSMVD RLWCSEKGYA
101 LHQHENGPFLL DVKRVQQILL RSPYIKVRAL WPSGDIPEDP SQAAYVLLLSL
151 WFFFSSVDVE ALLPSPQKEKE GKYIDPVLPK LSRIERVSL VFLSAFTLDD
201 LNEQGVNPLM NNEEFLFFIN KKAREHGIQD LKHEIMSSLE KTGVPFLDPSM
251 SFQVSQAMFS VYRYLRQRDL TTSELRCFHL LSCFKGDVVH CLASFENPKD
301 LADSDFLEAC KNVEWGEFIS ACEKALLKNP QGISIKDLKQ FLVR*

```

The cp6395 nucleotide sequence <SEQ ID 218> is:

```

1  ATGGAGAATG CTATGTCATC ATCGTTTGTG TATAATGGGC CTTCGTGGAT
51  TTTAAAAACG TCAGTAGCTC AGGAGGTATT TAAAAAGCAC GGTAAGGGGA
101 TTCAGGTTCT CTTAAGTACT TCAGTGATGC TTTTATAGG TCTTGGAGTC
151 TGTGCCTTTA TATTTCTCTA ATATCTGATT GTTTTGTGTT TGAATATAGC
201 TTTGCTTATG CTCGCTATAA GCTTGGTATT GTTCTCTCTA ATACGTCTCG
251 TACGCTCTTC AATGGTAGAT CGTTTGTGGT GTTCTGAAAA AGGATATGCT
301 CTTTCATCAAC ATGAGAACGG GCCTTTTTTG GATGTGAAGC GTGTACAGCA
351 AATTCCTCTA AGATCACCCCT ATATTAAGT TCGGGCTTTA TGGCCGTCTG
401 GAGATATCCC TGAGGATCCT TCACAAGCTG CGGTCTATT ACTTCTCTCT
451 TGGACTTTCT TTTTCATCCG GTGATGATAG GCTTTATTAC CGAGTCTCTA
501 AGAAAAGGAG GGTAAAGTATA TAGATCCTGT GCTGCCTAAG TTGTCTAGGA
551 TAGAGAGAGT CTCACCTTTA GTGTTTTTGA GTGCATTTAC TTTGGATGAC
601 TTAAACGAAC AGGGAGTCAA TCCTTTGATG AATAATGAGG AATTTTATT
651 TTTTATAAAT AAGAAAGCGC GTGAGCATGG GATTGAGGAT TTAAACACG
701 AGATTATGTC TTCGTTAGAG AAAACAGGAG TGCCATTAGA CCCCTCAATG
751 AGTTTTCAAG TTTCACAAGC GATGTTTTCT GTATATCGCT ACTTGAGACA
801 AAGGGATTTA ACGACTTCAG AATTAAGATG TTTTCACCTC TTAAGTTGTT
851 TTAAAGGGGA TGTGGTTTAT TGTTAGCTT CATTTGAAAA CCCTAAAGAT
901 TTAGCAGATT CTGACTTTT AGAAGCTTGT AAGAACGTGG AATGGGGTGA
951 GTTTATTTTC GCATGTGAGA AGGCTCTTTT AAAGAATCCG CAAGGAATTT
1001 CCATTAAGGA TCTAAACAA TTTTAGTGA GGTA

```

The PSORT algorithm predicts inner membrane (0.6307).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 109A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 109B) and for FACS analysis.

These experiments show that cp6395 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 110

The following *C.pneumoniae* protein (PID 4376396) was expressed <SEQ ID 219; cp6396>:

```

1  MIEFAFVPHT SVTADRIEDR MACRMNKLST LAITSLCVLI SSVCMIGIL
51  CISGTVGYA FVVGIIIFSVL ALVACVFFLY FFYFSSEEFK CASSQEFRFL
101 PIPAVVSALR SYEYISQDAI NDVIKDTMQL STLSSLLDPE AFFLEFPYFN
151 SLIVNHSMKE ADRLSREAFI ILLGEITWKD CETKILPWLK DPNITPDDFW
201 KLLKDHFDLK DFKRIATWI RKAYPEIRLP KKHCLDKSIY KGCKKFLLS

```

251 ENDVQYQRLL HKVCYFSGEF PAMVLGLGSE VPMVLGLPKV PKDLTWEMFM
301 ENMPVLLQSK REGHWKISLE DVASL*

The cp6396 nucleotide sequence <SEQ ID 220> is:

```

1  ATGATCGAGT TTGCTTTTGT TCCTCATACC TCCGTGACAG CGGATCGGAT
5  51  TGAGGATCGC ATGGCCTGTC GCATGAACAA GTTGTCTACT TTAGCAATTA
    101 CAAGTCTTTG TGTATTGATC AGTTCAGTTT GTATTATGAT TGGGATTTTA
    151 TGCATTTCTG GAACGGTTGG GACCTATGCA TTTGTTGTAG GAATTATTTT
    201 TTCTGTGCTT GCTTTGGTAG CATGTGTTTT CTTCTTTTAT TCTTTTATT
10  251 TTCTTCTGA GGAATTTAAG TGTGCTTCTT CGCAGGAGTT TCGTTTTTTG
    301 CCTATACCAG CTGTGGTTTC TGCATTGCGT TCCTATGAAT ACATTTCTCA
    351 GGACGCTATC AATGACGTTA TAAAAGATAC GATGCAGTTG TCTACCCTTT
    401 CTTCTCTTTT AGATCCCGAA GCTTTTTTCT TAGAATTTC TTATTTAAC
    451 TCTTTGATAG TGAATCATTC GATGAAGGAA GCGGATCGTT TGTCTCGAGA
15  501 GGCTTTTTTG ATTTTATTAG GTGAGATTAC TTGGAAGGAT TGTGAAACAA
    551 AAATTTTGCC ATGGTTGAAA GATCCTAATA TCACCTCTGA TGATTTCTGG
    601 AAGCTATTAA AAGACCATTT CGATTTAAAG GACTTTAAGA AGAGGATCGC
    651 CACTTGGAATA CGGAAGGCCT ATCCAGAAAT TAGATTACCG AAGAAGCAT
    701 GTTTAGATAA GTCTATCTAT AAGGGGTGTT GTAAGTTTTT ATTACTTTCT
20  751 GAGAATGATG TGCAATATCA GAGGTATTAT CATAAGGTCT GTTATTCTC
    801 TGGGGAGTTT CCTGCCATGG TTTTAGGTTT GGGAGTGAA GTGCCTATGG
    851 TGTTAGGACT CCCTAAGGTT CCCAAGGATC TTACCTGGGA GATGTTTATG
    901 GAAAATATGC CTGTTCTTCT GCAAAGCAAA AGAGAGGGGC ATTGAAAAAT
    951 CTCCTTGAA GACGTAGCCT CTCTTTAA
  
```

The PSORT algorithm predicts inner membrane (0.6095).

- 25 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 110A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 110B) and for FACS analysis.

These experiments show that cp6396 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 111

The following *C.pneumoniae* protein (PID 4376408) was expressed <SEQ ID 221; cp6408>:

```

1  MNTSLKRPLK SHFDVVGSL RPEHLKKTRE SLKEGSISLD QLMQIEDIAI
51  QDLIKKQKAA GLSFITDGEF RRATWHYDFM WGFHGVGHRH ATEGVFFDGE
35  101 RAMIDDTYLT DKISVSHHPF VDHFKFVKAL EDEFTAKQT LPAPAQFLKQ
    151 MIFPNNIEVT RKFYPTNQEL IEDIVAGYRK VIRDLYDAGC RYLQLDDCTR
    201 GGLVDPVRCV WYGIDEKGLQ DLIQQYLLIN NLVIADRPDD LVVNLHVCRG
    251 NYHSKFFASG SYDFIAKPLF EQTNVDGYLL EFDHERSGDF SPLTFISGEK
    301 TVCLGLVTSK TPTLENKDEV IARIHQAADY LPLERLSLSP QCGFASCEIG
    351 NKLTEEEQWA KVALVKEISE EVWK*
  
```

40 The cp6408 nucleotide sequence <SEQ ID 222> is:

```

1  ATGAATACTT CACTAAAAAG ACCTCTGAAA TCTCATTTTG ATGTTGTCGG
51  TAGTTTTTTG CGTCCTGAGC ATTTAAAAAA AACTAGAGAA AGCCTTAAAG
101 AAGGCTCTAT TTCTCTAGAT CAACTCATGC AAATTGAGGA TATCGCTATC
45  151 CAAGATTGTA TCAAAAAACA AAAAGCAGCA GGCTTTTCTT TTATTACTGA
    201 TGGAGAATTC CGCAGAGCTA CGTGCCATTA CGACTTCATG TGGGGTTTTT
    251 ATGGCGTAGG TCACCACAGA GCTACAGAAG GAGTTTTTCT TGATGGAGAA
    301 CGCGCTATGA TCGATGATAC CTATCTGACA GACAAGATCT CTGTATCTCA
    351 CCACCCATTT GTGGATCACT TTAAATTTGT AAAAGCTCTA GAAGATGAAT
50  401 TTACGACTGC AAAGCAAACT CTTCTGACAC CGGCACAGTT TTAAAGCAG
    451 ATGATCTTCC CTAATAATAT AGAGGTCACA CGTAAATTTCT ATCCTACAAA
    501 TCAGGAGCTA ATTGAAGATA TTGTTGCAAG TTATCGTAAA GTCATTTCGG
    551 ATCTTTATGA TGCTGGCTGC CGCTATCTCC AATTAGATGA CTGTACTCGG
    601 GGAGTTTAG TAGACCTCG AGTCTGTTTC TGGTATGGTA TCGATGAAAA
65  651 AGGTCTTCAA GATCTGATTC AACAATATCT TCTGATTAAT AATCTTGTA
    701 TTGCAGATCG TCCCAGATGAT CTAGTCGTTA ATTTACATGT ATGCCGTGGG
  
```

5 751 AACTACCACT CAAAATTCCTT TGCTAGTGGT AGTTATGACT TTATTGCAAA
 801 GCCCCTATTC GAACAAACAA ATGTAGACGG CTACTATTTA GAGTTTGATC
 851 ATGAGCGTTC TGGAGACTTC TCTCCTCTCA CCTTCATTTC TGGAGAAAAA
 901 ACTGTCTGCT TAGGTCTTGT TACCAGCAAA ACCCCTACAC TTGAAAATAA
 951 GGATGAGGTC ATTGCTCGCA TACATCAAGC AGCAGACTAC CTGCCCTTGG
 1001 AAAGACTCTC TCTAAGTCCA CAGTGTGGTT TTGCTTCATG TGAAATAGGA
 1051 AATAAATTAA CAGAAGAAGA GCAATGGGCT AAAGTTGCTC TAGTAAAGA
 1101 AATTTCGAA GAAGTTTGA AATAA

The PSORT algorithm predicts cytoplasm (0.2171).

- 10 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 111A) and also as a his-tagged product. The his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 111B) and for FACS analysis.

These experiments show that cp6408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 112

The following *C.pneumoniae* protein (PID 4376430) was expressed <SEQ ID 223; cp6430>:

1 MKLYSISSDV DTPWIFQLMS KVDSYFLGG NRIKVVSIVM QEPNLIIGKV
 51 ENVRISTIVK ILKILSFLIF PLILIALALH YFLHAKYANH LLVSKILERA
 101 PQYVPIGRS GDTASHYKLT TLVPVSQKNL QAMGSNPLEV EAALRTTKPS
 151 FFCVPAKYRQ IISSHGIRF SLDLEQLADD INLDSVSWPT EYLNSTMDFC
 201 SKADKRVIQN VQNLRTGTIYI NSVGKRSLLK FMLQHLFIDG ITQENPEALP
 251 NNTSGRLTLF PSVRYIYSHF TPQNPTIWPO VFFRQGPLDE DRGGGFEEILE
 301 QLQELGVRRP ICPSQGPDPN NFQGFQGI RI YWEDSYQPNK EV*

The cp6430 nucleotide sequence <SEQ ID 224> is:

25 1 ATGAAACTTT ATAGCATCTC TTCAGATGTA GATACACCTT GGATATTTCA
 51 GCTTATGTCA AAGGTAGATT CTTATCTTTT CTTAGGCGGG AATAGAATCA
 101 AGGTTGTATC TATAGTTATG CAAGAACCCTA ACTTAATTAT TGGAAAAGTA
 151 GAAAACGTTT GGATCTCCAC AATAGTGAAA ATATTAAAGA TTTTATCCTT
 201 CTTAATCTTC CCTCTGATTT TAATCGCTTT AGCCCTACAC TATTTTCTAC
 251 ATGCTAAATA TGCTAATCAC TTACTTGTAT CTAAGATTTT AGAAAGAGCT
 301 CCTCAGTATG TGCCATTTCC TGGTCGTTCA GGAGACACGG CGTCTCATT
 351 TAAATTAACA ACATTGGTTC CAGTATCCCA AAAAACTCA CAAGCTATGG
 401 GATCAAATCC TCTAGAAGTT GAAGCGGCTC TTCGAACCTAC AAAACCCCTCT
 451 TTTTCTGTG TACCTGCAAA ATACCGTCAG ATTATAAATT CAAGTCACGG
 501 CATTCGCTTT TCTTTAGATC TTGAACAACCT TGCTGATGAC ATTAATTTAG
 551 ATTTCGTTTC CTGGCCTACG GAGTATCTTA ACTCTACTAT GGATTTTTTGC
 601 AGCAAGGCAG ATAAACGTGT TATACAGAAAT GTACAAAATC TCGCGACAGG
 651 AACTTACATA AATTCTGTAG GAAAGCGTAG CCTTTTAAAA TTCATGTTAC
 701 AGCACCTATT TATTGATGGG ATCACACAAG AAAACCCCTGA AGCCCTTCCT
 751 AACAATACAT CTGGAAGACT GACTCTATTC CTTAGTGTTC GTTATATCTA
 801 TTCTCATTTT ACTCCACAAA ATCCTACAAT ATGGCCGCAA GTCTTTTTC
 851 GACAAGGTCC TCTAGATGAA GATCGAGGAG GAGGATTGTA GATCTTAGAG
 901 CAATTACAAG AGTTAGGAGT TAGGTTTCCA ATTTGCCCTC CTCAAGGACC
 951 AGACAATCCT AATTTTCAAG GTTTTCAAGG GATTCGTATC TATTGGGAAG
 45 1001 ATTCCTATCA ACCCAATAAG GAGGTTTAA

The PSORT algorithm predicts inner membrane (0.5140).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 112A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 112B) and for FACS analysis.

- 50 These experiments show that cp6430 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 113

The following *C.pneumoniae* protein (PID 4376439) was expressed <SEQ ID 225; cp6439>:

```

1  MSYDTLFLKNL EKEDSVHKIC NEIFALVPRL NTIACTEAII KNLPKADIHV
51  HLPGTITPQL AWILGVKNGF LKWSYNSWTN HRLLSFKNPH KQYSNIFRNF
101 QDICHEKDPD LSVLQYNILN YDFNSFDRVM ATVQGHFRFP GGIQNEEDLL
151 LIFNNYLQCC LDDTIVYTEV QQNIRLAHVL YPSLPEKHAR MKFYQILYRA
201 SQTFSKHGIT LRFLNCFNKT FAPQINTQEP AQEAVQWLQE VDSFTFGLFV
251 GIQSAGSESA PGACPKRLAS GYRNAYDSGF GCEAHAGEGI ETRTIFSSAK
301 VNPEGLIEIT RVTFSSLKRK QPSSLPIRVT CQLG*

```

10 The cp6439 nucleotide sequence <SEQ ID 226> is:

```

1  ATGTCTTATG ATACGTTATT CAAGAATCTT GAAAAGGAAG ATTCTGTACA
51  TAAGATATGC AATGAGATCT TTGCATTAGT ACCACGACTC AATACAATCG
101 CTTGCACCGA AGCTATCATC AAAAACCTCC CCAAAGCAGA TATCCATGTA
151 CACCTTCCTG GGACCATAAC ACCTCAATTA GCTTGGATT TAGGTGTGAA
201 AAATGGGTTC TTAATATGGT CTTATAATTC TTGGACCAAT CATCGATTAC
251 TTTCTCCTAA GAATCCTCAT AAACAATACT CCAATATTT CCGAAACTTT
301 CAAGATATCT GTCACGAAAA GGATCCGGAT TTAAGTGTAT TACAATATAA
351 TATCTTAAAT TACGATTTTA ATAGCTTTGA TAGAGTGATG GCTACAGTAC
401 AAGGACATCG CTTTCCTCCT GGAGGAATCC AAAATGAAGA AGACCTTCTT
451 CTCATTTTCA ATAACATCTT CCAGCAATGT CTGGACGATA CTATCGTGTA
501 TACTGAAGTA CAACAAAATA TCCGCCTTGC CCATGTTTTG TATCCTTCAT
551 TACCTGAAAA GCACGCGCGT ATGAAGTTT ATCAAACTTT GTATCGTGCT
601 TCGCAAACGT TTTCAAAACA CGGGATTACT TTACGATTTT TAACTGCTT
651 CAATAAACA TTTGCTCCAC AAATAAACAC ACAAGAACCT GCCCAAGAAG
701 CTGTTCAATG GCTCCAAGAG GTTGATTCTA CATTTCTTGG TCTATTTGTA
751 GGGATACAAT CCGCAGGATC AGAATCTGCG CCCGGAGCCT GTCCTAAGCG
801 ATTAGCTTCT GGATATAGAA ATGCTTATGA CTCAGGGTTT GGTGTGTAAG
851 CTCATGCTGG AGAAGGCATA GAGACCCGGA CTATTTTTC GTCAGCTAAG
901 GTAAATCCAG AGGGATTGAT CGAGATAACC CGAGTGACTT TCTCGTCTCT
951 TAAACGAAAA CAGCCATCTA GTTTACCCAT AAGAGTTACT TGCCAGTTAG
1001 GATAA

```

The PSORT algorithm predicts cytoplasm (0.1628).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 113A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 113B) and for FACS analysis.

These experiments show that cp6439 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 114

The following *C.pneumoniae* protein (PID 4376440) was expressed <SEQ ID 227; cp6440>:

```

1  LQSARRHLNT IFILDFGSQY TYVLAKQVRK LFVYCEVLPW NISVQCLKER
51  APLGIILSGG PHSVYENKAP HLDPEIYKLG IPILAICYGM QLMARDFGGT
101 VSPGVGEFGY TPIHLYPCEL FKHIWDCESL DTEIRMSHRD HVTTIPEGFN
151 VIASTSQCSI SGIENTKQRL YGLQPHPEVS DSTPTGNKIL ETVFQEICSA
201 PTLWNPLYIQ QDLVSKIQDT VIEVFDEVAQ SLDVQWLAQG TIYSDVIESS
251 RSGHASEVIK SHHNVGGLPK NLKLKLVLEPL RYLFKDEVRI LGEALGLSSY
301 LLDRHFFPGP GLTIRVIGET LPEYLAILRR ADLIFIEBLR KAKLYDKISQ
351 AFALFLPIKS VSVKGDERSY GYTIALRAVE STDFMTGRWA YLPDVLSSC
401 SSRIINEIPE VSRVYDISD KPPATIEWE*

```

The cp6440 nucleotide sequence <SEQ ID 228> is:

```

1  TTGCAGAGTG CAAGGAGACA TTTGAACACC ATATTTATTC TAGATTTTGG
51  ATCTCAATAT ACTTATGTAT TAGCAAAGCA AGTGCAGGAG TTATTTGTAT
101 ATTGCGAAGT TCTTCCCTGG AATATCTCTG TGCAATGTTT AAAAGAAAGA
151 GCGCCTTTGG GGATCATTCT CTCAGGAGGT CCTCACTCTG TCTATGAAA

```


201 CAAGGCTCCA CATTTAGATC CTGAAATCTA TAAACTTGGC ATTCCAATTC
 251 TAGCTATTTC CTATGGCATG CAGCTTATGG CTAGAGATTT TGGAGGGACT
 301 GTAAGCCCTG GTGTAGGAGA ATTTGGATAT ACGCCCATCC ATCTGTATCC
 351 TTGTGAGCTC TTCAAACACA TCGTCGACTG CGAATCTCTA GACACAGAGA
 401 TTCGGATGAG CCATCGGGAT CATGTTACGA CAATTCCTGA AGGATTTAAT
 451 GTAATCGCAT CCACCTCACA ATGCTCGATC TCAGGAATAG AAAATACCAA
 501 ACAACGGTTG TACGGGCTGC AATTTTCATCC CGAGGTTTCT GACTCCACTC
 551 CAACGGGAAA TAAGATTCTA GAAACTTTTG TTCAAGAGAT CTGTTCTGCT
 601 CCCACACTAT GGAATCCCTT GTATATTCAG CAAGACCTTG TAAGTAAAT
 651 TCAAGATACC GTTATTGAAG TATTTGATGA AGTCGCTCAG TCATTAGACG
 701 TACAATGGTT AGCTCAAGGA ACCATCTACT CAGATGTTAT TGAGTCCTCA
 751 CGCTCTGGAC ATGCCTCCGA AGTAATAAAA TCACATCATA ATGTAGGGGG
 801 GCTTCCAAA AATCTTAAGC TGAAGTTAGT CGAGCCCTTA CGTTATTTAT
 851 TTAAAGATGA AGTTCGAATT TTAGGAGAAG CCTAGGACT TTCTAGCTAT
 901 CTCTTGGACA GGCATCCTTT TCCTGGACCT GGCTTGACAA TTCGTGTGAT
 951 TGGAGAGATC CTTCTGAAT ATCTAGCCAT TTACGACGG GCGGACCTCA
 1001 TCTTTATAGA AGAGCTTAGG AAAGCAAAAC TCTACGATAA AATAAGCCAA
 1051 GCCTTTGCTC TATTTCTTCC TATAAAATCA GTATCTGTAA AAGGAGATTG
 1101 TAGAAGCTAT GGTATACCA TAGCATACG TGCTGTAGAA TCTACAGATT
 1151 TCATGACAGG ACGATGGGCC TACCTTCCAT GCGATGTTCT CAGTTCTTGC
 1201 TCATCGCGAA TTATTAATGA AATACCGAG GTAAGCCGAG TGGTCTATGA
 1251 TATTTCTGAC AAGCCACCAG CAACTATAGA ATGGGAATAG

The PSORT algorithm predicts cytoplasm (0.0481).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 114A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 114B) and for FACS analysis.

These experiments show that cp6440 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 115

The following *C.pneumoniae* protein (PID 4376475) was expressed <SEQ ID 229; cp6475>:

1 MNTYTFSPIL QKSFSFLLE KLDSYFFFGG TRTQILVITP TNIRLAACKR
 51 GCKVSTIEKI IKILSFILLP LVIIAIFILRY FLHKKFDKQF LCIPKVISNE
 101 DEALLGSRPQ AVEKAVREIS PAFFSIPRKY QLIRIDTPKD DAPSILFPIG
 151 IEIILKDLCT DTLKQSNLFL KREMDFLGHP EEKALFDSIC SIEKDQEWMS
 201 LESKLLITH FLKYLFSVSGI BQLNPGFNPE NGRGYFSEIS TAKIHFHQHG
 251 RYGPPIRSSGP IMKEI*

The cp6475 nucleotide sequence <SEQ ID 230> is:

1 ATGAATACCT ATACCTTCTC TCCTACACTT CAGAAAAGCT TCAGCCTATT
 51 TCTTTTAGAA AAATTAGACT CTACTTTTCT CTTTGGAGGG ACTCGTACAC
 101 AAATCTTAGT CATCACACCA ACCAATATTA GATTAGCAGC TAAAAAAGA
 151 GGGTGTAAGG TTTCTACTAT AGAAAAGATA ATCAAGATCC TCTCTTTTAT
 201 CCTGCTGCCC CTAGTTATCA TTGCCCTTAT ACTTCGCTAT TTCTTACATA
 251 AGAAATTCGA TAAACAGTTC TTGTGTATCC CAAAAGTCAT TTCTAACGAA
 301 GACGAAGCTC TTCTTGATC TAGACCACAA GCAGTTGAAA AAGCAGTTCC
 351 AGAAATATCT CCAGCCTTCT TCTCTATACC AAGAAAATAC CAACTTATTA
 401 GAATCGACAC TCCTAAAGAT GACGCTCCCT CAATCCTTTT CCCTATAGGC
 451 ATAGAGATCA TTCTCAAAGA TTTATGTATT GATACACTCA AGCAATCTAA
 501 TCTTTTCCTT AAAAGAGAAA TGGATTCTCT AGGTCATCCA GAAGAAAAAG
 551 CATTATTCGA CTCGATATGT TCTATAGAAA AAGATCAAGA ATGGATGAGC
 601 TTGGAAAGTA AAAAAGTTT AATCAGCAC TTCCTAAAGT ATCTCTTTGT
 651 CTCTGGAATC GAACAACTAA ATCCAGGCTT TAACCCAGAG AATGGGCGGT
 701 GGTATTTTTC AGAAATAAGT ACAGCAAAGA TCCATTTTCA TCAGCACGGT
 751 CGATATGGGC CAATCCGTTC TTCGGGACCC ATCATGAAGG AAATATAA

The PSORT algorithm predicts inner membrane (0.5373).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 115A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 115B) and for FACS analysis.

These experiments show that cp6475 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 116

The following *C.pneumoniae* protein (PID 4376482) was expressed <SEQ ID 231; cp6482>:

```

1  MLVELEALKR EFAHLKDQKP TSDQBITSLY QCLDHLFVL LGLGQDKFLK
51  ATEDEDVLF E SQRAIDAWNA LLTKARDVLG LGDIGAIYQT IEFLGAYLSK
101 VNRRAFCIAS EIHFLLKTAIR DLNAYYLLDF RWPLCKIEEF VDWGNDCEVI
151 AKRKLCTFEK ETKELESLL REEHAMEKCS IQDLQRKLSL IIEELHVDVSL
201 FCFSKTPSQE EYQKDCLYQS RLRYLLLLLYE YTLCLKTSTD FQEQARAKEE
251 FIREKFSLL E LEKGIKQTK E LEFAIAKSKL ERGCLVMRKY EAAAKHSLSL
301 MFEETVKSP RKDTE*

```

The cp6482 nucleotide sequence <SEQ ID 232> is:

```

1  ATGCTAGTAG AGTTAGAGGC TCTTAAAGA GAGTTTGCGC ATTTAAAGA
51  CCAGAAGCCG ACAAGTGACC AAGAGATCAC TTCACCTTAT CAATGTTTGG
101 ATCATCTTGA ATTCGTTTGA CTCGGGCTGG GCCAGGACAA ATTTTAAAG
151 GCTACGGAAG ATGAAGATGT GCTTTTGGAG TCTCAAAAAG CAATCGATGC
201 GTGGAATGCT TATTTGACAA AAGCCAGAGA TGTTTGTAGG CTGGGGGACA
251 TAGGTGCTAT CTATCAGACT ATAGAATTCT TGGGTGCCTA TTTATCAAAA
301 GTGAATCGGA GGGCTTTTGG TATTGCTTCG GAGATACATT TTCTAAAAAC
351 AGCAATCCGA GATTGGAATG CATATTACCT GTTAGATTTT AGATGGCCTC
401 TTTGCAAGAT AGAAGAGTTT GTGGATTGGG GGAATGATTG TGTGAAATA
451 GCAAAGAGGA AGCTATGCAC TTTTGAAAAA GAAACCAAGG AGCTCAATGA
501 GAGCCTTCTT AGAGAGGAGC ATGCGATGGA GAAATGCTCG ATTCAAGATC
551 TGCAAGGAA ACTTAGCGAC ATTATTATTG AATTGCATGA TGTTTCTCTT
601 TTTTGTTTT CTAAGACTCC CAGTCAAGAG GAGTATCAAA AGGATTGTTT
651 GTATCAATCA CGATTGAGGT ACTTATTGTT GCTGTATGAG TATACATTGT
701 TATGTAAGAC ATCCACAGAT TTTCAGAGC AGGCTAGGGC TAAAGAGGAG
751 TTCATTAGGG AGAAATTCAG CCTTCTAGAG CTCGAAAAGG GAATAAAACA
801 AACTAAAGAG CTTGAGTTTG CAATTGCTAA AAGTAAGTTA GAACGGGGCT
851 GTTTAGTTAT GAGGAAGTAT GAAGCTGCCG CTAAACATAG TTTAGATTCT
901 ATGTTCTGAAG AAGAACTGT GAAGTCGCCG CGGAAAGACA CAGAATAA

```

The PSORT algorithm predicts cytoplasm (0.4607).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 116A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 116B) and for FACS analysis.

These experiments show that cp6482 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 117

The following *C.pneumoniae* protein (PID 4376486) was expressed <SEQ ID 233; cp6486>:

```

1  VVVVALFILG IFFLSGSLAF LVHTSCGVLL GAALPILCIG LVLLAVALIV
51  FLCHKHKTRQ DLDYYDQDLD SLVIHKKEIP NDISELRVTF EKLQNLQFPH
101 TKDFSLSQE LQKFINCME KWLLEDEV T KFLIVRDRFL ETRRNFTTFG
151 EQVKGIQSN FDLHEEKSSL YLELYRLRKD LQVLLNFFLL PPGILKVDYD
201 EIEAIGLFI RLTSRLDKLD VKAQERKKFI NEMSREFKEV EKAFDIVDRA
251 TKKLMDRAK ESPARLFMGR TESLLEMKN BEALKNQGLD PENLSHPFLF
301 SPYQQLILN YLNSEIVLHH YEFLISGTVT SGLTLEECEN RMRAASTGLN

```

351 ALLVRKLQFR GAIKSAYFEK LTEIEKELRS LQDVIKSLEL ELIHKIKDIV
401 TEET*

The cp6486 nucleotide sequence <SEQ ID 234> is:

```

5      1 GTGGTGGTTG TCGCTTTATT TATCCTTGGG ATTTTCTTTT TATCTGGTTC
      51 TCTTGCATTG CTTGTTCATA CGTCTTGCGG AGTTCTTTTA GGAGCGGCGC
     101 TTCCCACTACT TTGCATAGGT CTTGTTTTAT TGGCTGTAGC TCTTATTGTT
     151 TTCTTATGTC ACAACACAAA GACTCGTCAA GATTAGATT ATTATGATCA
     201 AGATTTAGAT TCTTTGGTGA TTCATAAGAA AGAGATCCCC AATGACATCT
     251 CTGAGTTGCG GGTAAACATT GAAAAGTTGC AAAATCTGTT TCAGTTCCAT
    301 ACGAAAGATT TCTCTGATCT AAGCCAAGAG CTTCAGGGTA AATTATCAA
    351 TTGCATGGAG AAATGGCTAA CTTTGAAGA CGAAGTGACT AAATTTCTTA
    401 TTGTTGAGAG TAGATTTTTA GAAACCAGAA GAAATTTTAC CACTTTTGGG
    451 GAACAGGTTA AAGGATCCA AAGCAATATT TTTGATTGTC ATGAGGAAAA
    501 GTCTTCATTA TATTAGAAT TGTATAGGCT TAGGAAAGAC CTCCAAGTTC
    551 TATTAATAATT TTTTCTGCTC CCCCAGGTA TACTCAAGGT AGATTATGAT
    601 GAAATTGAGG CTATCAAAGG TCTGTTTATA AGATTAACCT CTAGATTAGA
    651 TAAGCTTGAT GTGAAAGCTC AGGAACGTAA GAAGTTCATT AATGAAATGA
    701 GTAGGGAATT TAAAGAAGTA GAGAAAGCTT TTGATATTGT CGATAGGGCA
    751 ACAAAAAGC TTATGGATAG AGCCAAGAAA GAAAGTCCGG CACGCTTTT
    801 CATGGGTAGA ACTGAGTCTC TCTTAGAAAT GAAAAAAAT GAAGAAGCCC
    851 TTAAAAATCA GGGCTAGAT CCTGAAATC TTTCCCATCC TGAACCTTTT
    901 AGTCCGTATC AACAGCTTTT AATTTTGAAT TATTTAATA GCGAAATAGT
    951 TCTGCATCAT TATGAGTTCC TTATTCTGG AACAGTAACT TCTGGCCTAA
   1001 CTCTTGAAGA ATGTGAAAT CGAATGAGG CGGCTTCTAC TGGGTTGAAC
   1051 GCCCTTCTGG TCGTAAGCT CCAGTTCAGA GGTGCTATAA AATCTGCGTA
   1101 TTTTGAAAAA CTCACAGAGA TTGAAAAAGA GTTACGATCA CTTCAAGACG
   1151 TAATAAAGTC ATTGGAAC TAAGATATAA AGATATAGTG
   1201 ACAGAAGAAA CTTAG

```

The PSORT algorithm predicts inner membrane (0.7474).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 117A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 117B) and for FACS analysis.

These experiments show that cp6486 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 118

The following *C.pneumoniae* protein (PID 4376526) was expressed <SEQ ID 235; cp6526>:

```

      1 MSPFKKIVNR LLCYISFQKE SRTLPIIIRE PRMTTKSLGS FNSVISKNKI
     51 HFISLGCERN LVDSEVMLGI LLKAGYESTN EIEDADYLIL NTCAFLKSAR
    101 DEAKDYLDHL IDVKKENAKI IVTGCMTSNH KDELKPWMSH IHYLLGSGDV
    151 ENILSAIESR ESGEKISAKS YIEMGEVPRQ LSTPKHYAYL KVAEGCRKRC
    201 AFCIIPSIKG KLRSKPLDQI LKEFRILVNK SVKEIILIAQ DLGDYGRDLS
    251 TDRSSQLESL LHELLKEPGD YWLRMLYLYP DEVSDGIIDL MQSNPKLLPY
    301 VDIPLQHIND RILKQMRRTT SREQILGFLE KLRKVVPQVY IRSSVIVGFP
    351 GETQEEFQEL ADFIGEGWID NLGIFLYSQE ANTPAAELPD QIPEKVKEER
    401 LKILSQIQKR NVDKHNQKLI GEKIEAVIDN YHPETNLLT ARFYGQAPV
    451 DPCIIVNEAK LVSHFGERCF IEITGTAGYD LVGRVVRKSQ NQALLKTSKA
    501 *

```

The cp6526 nucleotide sequence <SEQ ID 236> is:

```

50      1 ATGAGTCCTT TTAAGAAAAT AGTAAATCGC TTAATATGCT ATATTTCTTT
     51 TCAAAAAGAA TCAAGAACTC TCCCAATCAT TATTAGAGAA CCTAGGATGA
    101 CAACAAAAG TTTAGGATCT TTCAATTCAG TTATTTCCAA AAATAAAATT
    151 CATTTTATTA GTTTGGGATG CTCTCGGAAC CTTGTAGATA GCGAAGTCAT
    201 GCTAGGCATT CTTCTTAAGG CAGGTTACGA GTCTACTAAT GAAATTGAAG
    251 ATGCTGACTA TTTAATTTTA AATACCTGTG CGTTTTTAAA AAGTGCTAGA
    301 GATGAAGCTA AAGATTATCT AGACCATCTA ATTGATGTAA AAAAAGAGAA

```

5
10
15
20

```

351 CGCTAAATTT ATTGTAAGT GATGCATGAC TTCCAACCAC AAAGATGAGC
401 TTAAACCCTG GATGTCACAC ATCCATTACC TACTAGGTTT TGGGGATGTT
451 GAGAATATTC TTTCGTCTAT TGAGTCTCGT GAATCTGGAG AAAAATCTC
501 TGCAAAGAGT TACATTGAGA TGGGAGAAGT TCCAAGACAG CTTTCCACAC
551 CAAAACACTA TGCCATATTTA AAAGTTGCTG AGGGCTGTAG AAAACGTTGT
601 GCTTTTGTGA TTATTCCTTC CATTAAAGGA AAGCTCCGCA GCAAACCTCT
651 GGATCAAATTT CTTAAAGAAT TCCGCATCCT TGTAACAAG AGTGTGAAAG
701 AGATTATATT GATAGCTCAA GACCTAGGAG ATTATGGAAA GGATCTCTCT
751 ACAGACCGCA GTTCGCAGCT AGAATCACTA TTACATGAGT TACTGAAAGA
801 GCCTGGTGAT TATTGGCTGC GGATGTTGTA TTTATATCCT GATGAAGTGA
851 GTGATGGCAT TATAGATCTT ATGCAATCTA ATCCCAAAC TCTTCCCTAT
901 GTAGATATTC CTTACAGCA CATTAAACGAC CGTATTTTAA AGCAAATGCG
951 AAGAACGACT TCTAGGGAGC AAATCCTAGG ATTCTTAGAA AAATTACGTG
1001 CCAAGGTTCC TCAGGTCTAT ATCCGTTCTT CTGTTATTGT GGGTTTCCCG
1051 GGTGAAACTC AGGAAGAATT CCAGGAGTTA GCTGATTTTA TTGGTGAGGG
1101 TTGGATTGAT AATCTCGGAA TTTTCTTGTA CTCTCAAGAA GCGAATACCC
1151 CCGCAGCAGA ACTCCCTGAC CAGATACCAG AAAAAGTTAA AGAATCGAGG
1201 TTGAAAATTC TATCTCAAAT TCAGAAACGC AATGTGGATA AACATAATCA
1251 GAAGCTCATTT GGGGAAAAAA TAGAAGCAGT TATTGATAAC TATCATCTCTG
1301 AAACGAATCT TTTACTCACT GCAAGGTTCT ATGGACAAGC TCCTGAAGTG
1351 GACCTTGTGA TTATTGTAAA TGAGGCGAAG CTTGTTTCTC ATTTTGGAGA
1401 AAGATGCTTT ATAGAAATCA CAGGGACTGC TGGTTACGAC CTTGTAGGCG
1451 GTGTGTGAAA AAAATCTCAG AACCAAGCTT TGCTAAAAAC TAGCAAAGCT
1501 TAG

```

25 The PSORT algorithm predicts cytoplasm (0.1296).

The protein was expressed in *E. coli* and purified as a GST-fusion product (Figure 118A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 118B) and for FACS analysis.

30 These experiments show that cp6526 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 119

The following *C. pneumoniae* protein (PID 4376528) was expressed <SEQ ID 237; cp6528>:

35

```

1 MKNNINNNEC YFKLDSTVDG DLLAANLKTf DTQAQGISST ETFSVQGNAT
51 PKDQVSATGL TSGTTYNLNA QNFTSSQISI DFKNNRLSNC ALPKEDCDPV
101 PANYVRSPEY FFCCKPLIGD FDFNSGESYL PLTGSEYTLY QSRNVNSIFR
151 FIGWKQSTRE LTVGGNTAIQ FLAAGTYIVS FTVGKRWGWN NGWGGAIYIN
201 NGLGQVQCES TIYSGGYAT IGTGTSIYR ASVDVAPNPN DPNASDRYRA
251 GIFYLSNGGS SAGIGNYSFS LLYYPDDR*

```

The cp6528 nucleotide sequence <SEQ ID 238> is:

40
45
50
55

```

1 ATGAAAAACA ATATTAATAA TAATGAGTGC TATTTTAAAT TAGACTCAAC
51 TGTAGATGGT GATTTGTTAG CAGCCAATCT CAAGACCTTT GATACACAGG
101 CCCAAGGAAT CTCATCGACT GAAACATTTT CTGTTCCAGG GAATGCAACA
151 TTAAAGATC AAGTTTCAGC AACTGGATTA ACTTCAGGAA CTACTATATA
201 TTAAATGCA CAAACTTTA CTTCCTCCCA AATCTCTATA GATTTTAAAA
251 ATAATCGTCT GAGTAATTGT GCATTGCCAA AAGAAGACTG CGATCCGGTG
301 CCAGCGAATT ATGTTCTGTT TCCCGAATAT TTTTCTGTT CCAAGCCTCT
351 GATCGGAGAT TTTGATTTTA ACTCAGGGGA ATCTTATTTG CCTCTGACTG
401 GTTCGGAATA TACTCTATAT CAGTCACGTA ATGTAATAG TATATTTCTG
451 TTTATAGGAT GGAAGCAAAG TACACGAGAA TTAAGTGTAG GGGGAAATAC
501 TGCGATACAA TTTCTTGCAG CAGGAACCTA TATCGTTTCA TTTACTGTTG
551 GTAAACGCTG GGGATGGAAT AATGGTTGGG GAGGAGCCAT TTATATCAAT
601 AATGGTTTAG GACAAGTCCA ATGTGAAAGC ACGATTATA GTGGTGGAGG
651 GTATGCAACA ATAGGTACAC TGGGGACCTC AATATATAGA GCCTCTGTAG
701 ATGTAGCTCC TAATCTTAAT GATCCGAATG CTTCCGATCG CTATAGAGCG
751 GGTAATTTCT ATCTCAGTAA CGGTGGTTCT AGTGCAGGTA TAGGGAATTA
801 CTCCTTTTCT CTTCTCTATT ATCCGGACGA TAGAGGGTAG

```

The PSORT algorithm predicts cytoplasm (0.1668).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 119A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 119B) and for FACS analysis.

- 5 These experiments show that cp6528 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 120

The following *C.pneumoniae* protein (PID 4376627) was expressed <SEQ ID 239; cp6627>:

```

10      1  MKCSPLTLVP  HIFLKNDEC  HRSCSLKIRT  IARLILGLVL  ALVSALSFVF
      51  LAAPISYAIG  GTLALAAIVI  LIITLVVALL  AKSKVLPIN  ELQKIYNRY
      101  PKEVYFVKI  HSLTVNELKI  FINCWKSGTD  LPPNLHKAE  AFGIDILKSI
      151  DLTLFPEFEE  ILLQNCPLYW  LSHFIDKTES  VAGEIQLNKT  QKVYGLLGPL
      201  AFHKGYTIF  HSYTRPLLTL  ISESQYKFLY  SKASKNQWDS  PSVKKTCEEI
      251  FKELPHNMIF  RKDVQGISQF  LFLFFSHGIT  WEQAQMIQLI  NPDNWKMLCQ
      15  301  FDKAGGHCMS  ATFGGFLNTE  TNMPDPVSSN  YEPTVNFMTW  KELKVLLEKV
      351  KESPMHPASA  LVQKICVNTT  HHQNLKRWQ  FVRNTSSQWT  SSLPQYAFHA
      401  QTYKLEKKIE  SSLPIRSSL*

```

The cp6627 nucleotide sequence <SEQ ID 240> is:

```

20      1  ATGAAGTGTA  GTCCTTTAAC  ACTAGTTCCC  CATATATTTT  TAAAAATGA
      51  CTGCGAATGT  CATAGATCTT  GTTCTTTAAA  AATTAGGACA  ATTGCCCGAC
      101  TCATTCTTGG  GCTTGTCTTA  GCTCTTGTTA  GCGCACTTTC  TTTTGTTTTC
      151  CTTGCTGCGC  CGATTAGCTA  TGCTATTGGA  GGAACCTTAG  CTTTAGCCGC
      201  TATCGTAATC  TTGATTATAA  CGCTAGTCGT  AGCACTGCTA  GCTAAATCAA
      25  251  AGGTTCTGCC  CATCCCCAAC  GAACTTCAGA  AGATTATTTA  CAATCGCTAT
      301  CCTAAAGAAG  TCTTTTATTT  CGTGAAAACA  CACTCCCTGA  CTGTTAACGA
      351  ATTAAAAATA  TTTATTAATT  GCTGAAAAG  CGGTACAGAC  CTGCCCTCCGA
      401  ATTTACATAA  AAAAGCAGAG  GCTTTCGGGA  TCGATATTTT  AAAATCTATA
      451  GATTTAACCC  TGTTCACAGA  GTTCGAAGAG  ATTCTTCTTC  AAAACTGCCC
      501  GTTATACTGG  CTCTCCCAT  TTATAGACAA  AACTGAATCT  GTTGCTGGGG
      30  551  AAATCGGATT  AAATAAAACA  CAAAAGTTT  ATGGTTTACT  TGGGCCCTTA
      601  GCGTTTCATA  AAGGATATAC  AACTATTTTC  CACTCTTATA  CACGCCCTCT
      651  ACTAACATTA  ATCTCAGAA  CACAGTATAA  GTTCCTATAT  AGTAAAGCGT
      701  CTAAGAAATCA  ATGGGATCT  CTTCTGTGA  AAAAAACCTG  CGAAGAAATA
      751  TTCAAGGAAC  TCCCCACAA  TATGATTTTC  CGGAAGGATG  TTCAAGGAAT
      35  801  CTCACAAATC  TTATTTCTTT  TCTTTCTCA  TGGTATCACT  TGGGAACAGG
      851  CTCAGATGAT  TCAACTTATA  AATCCTGATA  ATTGGAAAAT  GTTGTGTGTC
      901  TTTGATAAAG  CAGGAGGCCA  CTGTCCATG  GCAACATTG  GAGGCTTTTT
      951  GAATACTGAA  ACAAATATGT  TCGATCCAGT  ATCCTCTAAC  TATGAACCTA
      1001  CAGTGAACCT  CATGACGTGG  AAAGAATTGA  AGGTTTTACT  AGAGAAAGTA
      40  1051  AAAGAAAGTC  CTATGCACCC  AGCGAGTGCT  CTTGTTCAGA  AGATATGCGT
      1101  AAATACAACG  CACCATCAAA  ATCTGTAAA  ACGATGGCAA  TTTGTTGCGT
      1151  ATACGAGTTC  ACAATGGACA  TCAAGCTTAC  CTCAGTATGC  TTTCCACGCC
      1201  CAAACCTACA  AACTAGAGAA  AAAAAAGAA  AGCAGTCTCC  CTATACGATC
      1251  TTCCCTATAA

```

- 45 The PSORT algorithm predicts inner membrane (0.7198).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 120A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 120B) and for FACS analysis.

- 50 These experiments show that cp6627 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 121

The following *C.pneumoniae* protein (PID 4376629) was expressed <SEQ ID 241; cp6629>:

```

1 MSNITSFVIQ NNRSNYYFE LKNSTTIHIV ISAILLCGAL IAFLCVAAPV
51 SYILSGALLG LGLLIALIGV ILGIKKITPM ISSKEQVFPQ ELVNRIRAHY
101 PKFVSDFVSE AKPNLKDLS FIDLLNQLHS EVGSSTNYNV SEELQOKIDT
151 FEGIARLKNE VRTASLKRLE SAASSRPLFP SLPKILQKVF PFFWLGEFIS
201 AGSKVVELHR VKKIGGSLEE DLSDYIKPEM LPTYWLIPLD FRPTNSSIILN
251 LHTLVLARVL TRDVFQHLKY AALNGEWNLN HSDLNMTKQQ LFAKYHAAVQ
301 SYKHLSPQSL QEDEFYNLLI CIFKHYSWK QMSLIKTVEA DLWENLCCLT
351 LDHTGRPQDM EFASLIGTLY TQGLIHKESE AFLSSLTLLS LDQFKTIRRQ
401 STNIAMFLEN LATHNSTFRS LPPIVHPLK RSVFSQPEED ESSLLIG*
```

The cp6629 nucleotide sequence <SEQ ID 242> is:

```

1 ATGAGTAATA TAACCTCGCC AGTTATTCAA AATAATCGCT CTGTGAATTA
51 TTATTTTGAA TTAAAGAATT CAACCACTAT TCATATTGTT ATCAGTGCCA
151 TCTTACTCTG CGGAGCTTTG ATAGCTTTCT TGTGTGTAGC AGCTCCTGTT
201 TCCTATATTC TAAGTGGCGC ATGTGTTAGGA TTAGGATTAT TAATAGCCTT
251 GATTGGTGTG ATTTTAGGAA TAAAAAAAT CACGCCTATG ATTTTCATCAA
301 AAGAACAAGT ATTCCCCCAA GAACTCGTAA ATAGAATCAG GCGCGACTAT
351 CCTAAATTG TCTCTGATTT TGTTCAGAA GCTAAACCAA ATCTTAAAGA
401 CATCTACAAA TTACAACGTA TCTGAAGAAC TACAACAGAA AATAGATACG
451 TTCGAGGGTA TCGCACGCTT AAAAAATGAA GTCCGTACTG CTCTCTTAA
501 AAGACTTGAA AGCGCTGCTT CTCCCGTCC CCTCTTCCCC TCTTTACCAA
551 AAATCTTACA AAAGGTATTT CCATTTTCTT GGTTAGGAGA GTTTATTTCT
601 GCAGGCAGCA AGGTTGTAGA GCTCCATCGA GTTAAGAAAA TTGGAGGCAG
651 CCTCGAAGAA GACCTTAGTG ATTATATAAA ACCAGAGATG CTTCCTACCT
701 ATTGGTTGAT TCCTTTTAGAT TTTAGACCAA CAAATTCTCT TATTCTAAAT
751 CTACACACAT TAGTTTTAGC TAGAGTCTTA ACTCGTGATG TTTTCAACA
801 TCTTAAGTAT GCAGCATTA ATGGCGAGTG GAACCTGAAT CATAGTGATC
851 TAAATACTAT GAAACAGCAG CTCTTTGCTA AATATCATGC GCGGTATCAA
901 TCCTATAAAC ATCTATCTCA ACCCTCTCTT CAAGAGGATG AATTCATAA
951 CCTGCTCTTG TGTATTTTGA AGCATAGGTA CTCGTGGAAG CAGATGTCCT
1001 TAATAAAAC AGTCCCGGCT GATTTATGGG AAAACCTCTG TTGCTTGACT
1051 TTAGACCATA CAGGACGACC CCAAGACATG GAATTTGCCT CTCTAATTGG
1101 TACTCTCTAC ACACAAGGCC TAATTCATAA AGAAAGCGAA GCATTCTTTT
1151 CTTCATTGAC ACTCCTTAGT TTAGATCAGT TTAACCGAT CCGTCGTCAG
1201 TCAACCAATA TAGCGATGTT CCTTGAGAA TTAGCAACTC ATAATTCCAC
1251 CTTTAGAAGC TTACCACCTA TAACAGTCCA TCCACTCAAG AGAAGCGTCT
1301 TCTCCAACC TGAAGAAGAC GAGTCCTCCC TGCTGATAGG TTAG
```

40 The PSORT algorithm predicts inner membrane (0.5776).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 121A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 121B) and for FACS analysis.

These experiments show that cp6629 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 122

The following *C.pneumoniae* protein (PID 4376732) was expressed <SEQ ID 243; cp6732>:

```

1 MEMMSPFQQP EQCHFVVVGS FLRPESLTRA RSDFEGRIV YEQMRVEDA
51 AIRNLIKQKT EAGLIFFTDG EFRYSWDFD FMWGFHGVDR RRDSNDPEIG
101 VYLKDKISVS KHPFIEHFEP VKTFEKGNAK AKQTIPSPSQ FFHEMIFAPN
151 LKNTRKFYPT NQELIDDIVF YYRQVIQDLY AAGCRNLQLD DCAWCRLLDI
201 RAPSWYGVDS HDRLQEILEQ FLWIHNLVMK DRPEDLFVSL HVCGRGDYQAE
251 FFSRRAYDSI EEPLFAKTDV DSYHYWALD DKYSGGAEPL AYVSGEKHVC
301 LGLISSNHSC IEDRDAVVSF IYEAASYIPL ERLSLSPQCG FASCEGDHRM
```

351 TEEEQWKIA FVKEIAKEIW G*

The cp6732 nucleotide sequence <SEQ ID 244> is:

```

1  ATGGAATGA  TGAGCCCAT  CCAACAACCT  GAGCAATGTC  ATTTTGATGT
5  51  TGTGGGAAGT  TTCTTACGTC  CTGAAAGTCT  TACACGAGCA  CGCTCTGATT
    101  TTGAAGAAGG  AAGAATTGTC  TATGAGCAGA  TGCAGATTGT  CGAAGATGCT
    151  GCTATTCGTA  ATCTCATAAA  AAAGCAAACA  GAAGCAGGTC  TTATCTTTTT
    201  TACTGATGGG  GAATTCGGTA  GGTATAGTTG  GGATTTTCGAC  TTTATGTGGG
    251  GATTCCATGG  CGTGGATCGT  CGCAGGGACT  CTAATGACCC  TGAATTTGGA
    301  GTGTATCTTA  AAGATAAAAT  CTCCGTATCA  AAACATCCGT  TTATAGAACA
10  351  TTTCGAGTTT  GTCAAAACCT  TTGAGAAGGG  AAATGCAAAA  GCAAAACAAA
    401  CGATTCCCTC  TCCATCACAA  TTTTTCATG  AGATGATTTT  TGCTCTTAAT
    451  CTGAAAATA  CTCGGAAGTT  TTATCCTACG  AATCAAGAGC  TAATTGATGA
    501  TATTGTCTTT  TATTATCGCC  AAGTCATCCA  AGATCTTTAT  GCTGCAGGTT
15  551  GTCGTAATTT  GCAGTTGGAC  GATTGTGCTT  GGTGTCGCCT  CTTGCTATATA
    601  CGAGCGCCTT  CTTGGTATGG  TGTTCATTCT  CATGACAGGT  TGCAGGAAAT
    651  TTTAGAACAG  TTTTATGGA  TCCATAATTT  AGTGATGAAG  GATAGACCCG
    701  AGGATCTTTT  TGTAAGTCTG  CATGCTGTGC  GTGGTGATTA  TCAGGCCGAG
    751  TTTTCTCTTA  GACGAGCTTA  TGATTCTATA  GAGGAGCCTT  TATTTGCTAA
    801  GACCGATGTG  GATAGTTATC  ACTATTATTG  GGCTCTTGAT  GATAAGTATT
20  851  CAGGAGGTGC  TGAGCCTTTA  GCTTACGTCT  CTGGAGAGAA  ACACGCTCTG
    901  TTGGGATTGA  TCTCCAGCAA  CCATTCTTGT  ATTGAAGATC  GAGATGCTGT
    951  GGTTCCTCGT  ATTTATGAAG  CTGCGAGCTA  CATTCCCTTA  GAGAGACTTT
1001  CTTTGAGCCC  GCAATGTGGG  TTTGCTTCTT  GTGAGGAGA  CCATAGAATG
1051  ATTGAAGAAG  AACAGTGGAA  GAAGATCGCC  TTTGTGAAAG  AGATTGCTAA
25  1101  AGAGATCTGG  GGATAA

```

The PSORT algorithm predicts cytoplasm (0.2196).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 122A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 122B) and for FACS analysis.

30 These experiments show that cp6732 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 123

The following *C.pneumoniae* protein (PID 4376738) was expressed <SEQ ID 245; cp6738>:

```

35 1  VWLRFLLLV  YDEKEKDVV  VCNHSEPNIL  GLPPEAVSQL  IEELSDEGYS
    51  YLNVVRCDLS  GETTVQQRLL  LNADEGRSMT  VVISELPEGH  PDIRNLQLAS
    101  ERIFVSREKE  AADAYASGCK  VVAFDDEHLP  WVSSHIAAYE  EIREKQEQTM
    151  QGSLTEEQLG  ALLCNTVSTE  KNLAFAALDAV  IKQSVWRFRN  PDLFAYEREA
    201  LEASVTDALV  SVSNLDMIP  YTSSQGIVIE  DSSIVRTSQE  HTLIVNCAAF
    251  DKLASQIEFL  CPSDVLPIG  KDPLISDDED  EELNPKVSSA  ADSKDKT*

```

40 The cp6738 nucleotide sequence <SEQ ID 246> is:

```

1  GTGTGGCTGC  GCTTTTACT  TTTAGTGTCC  TATGATGAGA  AGGAGAAAGA
51  CGTAGTTGTC  GTTTGTAATC  ATTCTGAACC  TAATATCCTC  GGCCTGCCTC
101  CTGAAGCAGT  CTCTCAGCTT  ATTGAAGAGC  TTAGCGATGA  AGGCTATAGC
151  TATCTGAATG  TAGTGCGTTG  TGATCTCTCC  GGGGAGACTA  CGGTTCAACA
45  201  ACGTCTGCTA  TTGAATGCCG  ATGAAGGGAG  ATCTATGACG  GTGGTGATCT
    251  CAGAGCTTCC  TGAAGGGCAC  CCCGATATTC  GGAATTTGCA  GTTGGCATCC
    301  GAAAGAATTT  TTGTTTCTCG  TGAAAAGAGAA  GCTGCTGATG  CCTATGCTTC
    351  AGGATGTAAA  GTGGTCGCTT  TCGATGATGA  GCATCTCCCT  TGGGTCTCCA
    401  GTCATATTGC  CTACGCGGAG  GAGATCAGAG  AGAAACAAGA  ACAAACAATG
50  451  CAAGGGTCTT  TAACTGAAGA  GCAGTTAGGA  GCACTCCTCT  GCAACACAGT
    501  CTCCACAGAG  AAAAATCTAG  CCTTTGCTCT  AGACGCCGTG  ATAAAACAGT
    551  CTGTGTGGAG  ATTCCGCAAT  CCGGATCTTT  TTGCTTATGA  GAGAGAAGCT
    601  CTAGAGGCTT  CAGTAACAGA  TGCTTTAGTA  TCTTACGTTT  CAAATTTAGA
    651  CATGATACCG  TACACAAGTT  CTCAGGCGAT  AGTCATAGAA  GATAGTAGTA
55  701  TCGTCCGTAC  CTCTCAAGAG  CATACTACTCA  TTGTGAAGTG  TGCAGCATTC

```


751 GATAAGTTAG CGAGCCAAAT AGAGTTCTTA TGCCCCAGTG ACGTGTGCGC
 801 CATTTCTGGT AAAGACCCTT TGATTCTGA TGATGAGGAT GAGGAACCTGA
 851 ATCCTAAAGT TTCATCTGCT GCAGACTCTA AAGATAAAAC CTAG

The PSORT algorithm predicts cytoplasm (0.1587).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 123A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 123B) and for FACS analysis.

These experiments show that cp6738 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

10 Example 124

The following *C.pneumoniae* protein (PID 4376739) was expressed <SEQ ID 247; cp6739>:

1 MTHCLHGWFS VVRHHFVQAF NFSRPLYSRI THFALGVIKA IPIVGHVLMG
 51 VDWLISHCFE RGVJHPGFPS DIAPILKVEK IAGRDHISRI ENQLKSLRKT
 101 IEVEDLDKVV GQYQENPYAD MASSEVLKLD KGVHVSSELGK AFSRVNRNRT
 151 RSYSYAPTPQ LDSIAIVGID LVSPPEEQENL VRLANEVIQL YPKSKTTLVL
 201 LIDFNKEWVG DISSDKEKQL RSLGLHSEVQ CLSVLEPQGA EGEDTKHFDL
 251 MVGCGYKDSY LREGKILQQA LGTSLGTVPW VNVMTLPSR YRSRLSLPIN
 301 TEKDKTELYK ELSRTHQLH TLGMGLGAQD SGLLLDRQRL HAPLSQGS HC
 351 HSYLADLTHE ELKILLFSAP VDAKNISKKE LREVSLNFAN DTSVECGCAF
 401 YF*

The cp6739 nucleotide sequence <SEQ ID 248> is:

1 ATGACTCATT GCTTACATGG TTGGTTTCT GTAGTTCGTC ATCACTTTGT
 51 GCAGGCGTTT AATTTCTCAC GTCCTTTATA TTCTCGAATT ACCCACTTCG
 101 CTTTAGGGGT GATTAAGGCC ATCCCCATTG TAGGGCATCT GTTATGGGA
 151 GTCGATTGGT TGATCTCTCA TTGCTTCGAG AGGGGAGTCT CACACCTTGG
 201 GTTCCCTTCA GATATTGCTC CTATACTGAA AGTAGAAAAG ATCGCGGGCC
 251 GAGATCATAT TTCTAGAATC GAAAATCAGC TAAAGAGCCT TAGGAAAAC
 301 ATCGAGGTTG AAGATCTAGA TAAAGTCCAC GGGCAATATC AAGAGAATCC
 351 TTATGCAGAT ATGGCCTCTA GTGAGGTCTT TAAACTCGAT AAGGGAGTTC
 401 ATGTTAGCGA GCTTGGCAAA GCCTTTTCTA GAGTTCGCAA TCGCATCACC
 451 AGATCCTATA GTTATGCCCC TACTCCTCAG TTGGACTCTA TAGCTATTGT
 501 TGGTATAGAT CTCGTCAGTC CTGAAGAACA AGAGAATTTA GTACGCTTGG
 551 CGAATGAGGT CATTCAACTC TATCCCAAAT CAAAGACAAC TCTATATCTT
 601 CTTATCGATT TTAATAAGGA GTGGGTAGGG GATATCTCCT CTGATAAGGA
 35 651 AAAACAGCTC CGTTCTCTAG GTCTACATTC TGAAGTTCAG TGTCTTTCCG
 701 TCTTGAACCC TCAGGGTGCC CAGGGCGAAG ATACGAAACA CTTTGACCTT
 751 ATGGTCGGCT GTTATGGGAA GGATTCTTAC TTAAGGGAGG GTAAAATTTT
 801 ACAGCAGGCC CTAGGGACTT CGTTAGGTAC TGTTCCCTGG GTGAATGTTA
 851 TGCACACATT GCCATCTAGG TATAGATCTC GGCTTTCCTT ACCTATAAAT
 40 901 ACCGAAAAGG ATAAGACAGA GCTTTATAAA GAGATTCTC GTACACACCA
 951 TCAGTTGCAT ACTTTGGGAA TGGGACTTGG AGCCCAGGAT TCAGGATTGC
 1001 TCTTAGACCG GCAACGACTC CATGCTCCTT TATCTCAAGG GTCTCACTGC
 1051 CATTCCATATC TTGCAGATCT CACCCATGAA GAGCTGAAAA TTTTGTATAT
 1101 TTCAGCATTG GTGGATGCTA AGAACATAAG TAAGAAAGAG CTTCTGTAGG
 45 1151 TATCTCTAAA TTTTGCTAAC GATACTTCCG TAGAGTGTGG CTGCGCTTTT
 1201 TACTTTTAG

The PSORT algorithm predicts inner membrane (0.2190).

- The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 124A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 124B) and for FACS analysis.

These experiments show that cp6739 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 125

The following *C.pneumoniae* protein (PID 4376741) was expressed <SEQ ID 249; cp6741>:

```

5      1 MASCLSAWFS IVREHFYRAF DFSLPFCARI TEFVLGVKIG IPVVGHIIVG
      51 IEWLVSRYLE SFVTKPTFVS DVVSLKTEK VAGRDHIARV VETLKRQRVA
101    VAPEDEDKVH GKIPVHPFGG IQPVEVLTLY PEVQDATLGL AFSKIRNRVR
151    QAYLQAPRPK LQKIYIIGND MNPFEVDLFL HLAFLCNETQ RLYPDATISL
10      201 YLTASGGRNA MDKKNRKLLS DCELNPKIAC LDFNQGDVVK QATCDCWMVY
251    HGENDQGTIN QIQEELEKSG EETFWIHVGQ KPLSQSLWDF SPFSSELMKG
301    DKEKALEYSE LEKEQLYSRL VVVGERSVSL SLGFGDSRSG ILMDPKRVHA
351    PLSEGHYCHS YLADLENPGL QKTLAALFIN PKELSSSTILQ PISLNLILNS
401    KTYLRQHFGE FERMSRSDRN VVVVCDSDWV GTDWKEEPSF QHFIMELECR
15      451 GYSHFNIFAF RSNMCMVEER RILNESSQEK AFTMIFCEDS VSQGDIRCLH
501    LASEGMLCGK ECVAVDVVYS GCANFMMEEV LTLERESNLW NRKHGLWKRE
551    VRKQKQEAAL DQDESEIYVC NQLTAQQNFA CS*
```

The cp6741 nucleotide sequence <SEQ ID 250> is:

```

      1 ATGGCTTCTT GTTATCTGCT CTGGTTTCTT ATAGTTCGTG AGCACTTTTA
20      51 TCGAGCCTTT GATTTTCTCT TGCCGTTTGT TGCTCGTATT ACGGAATTTG
101    TATTAGGGGT CATCAAGGGG ATCCCTGTTG TGGGTACATAT TATTGTTGGG
151    ATAGAGTGGC TCGTTTCTAG GTATTTAGAG AGTTTCGTGA CCAAGCCGAC
201    ATTTGTCTCT GATGTGGTGA GTCTTCTGAA AACAGAGAAA GTTGCTGGTC
251    GCGATCACAT TGCTCGTGTA GTGGAGACTT TGAAGAGGCA GAGAGTCGCT
301    GTGGCTCCTG AAGATGAGGA TAAGGTCCAT GGAAGATTG CTGTGCATCC
25      351 TTTCGGGGGA ATCCAACCTG TAGAAGTTCT CACTCTCTAT CCCGAAAGTC
401    AAGATGCAAC GTTAGGGCTT GCCTTCTCTA AAATTCGTAA TCGTGTAAAG
451    CAGGCGTATT TGCAAGCTCC ACGGCCAAAA CTGCAGAAAG TTTACATCAT
501    AGGAAACGAT ATGAATCCTT TTGAAGTTGA CGACTTCTTG CATCTAGCCC
551    GTCTCTGTAA TGAACCTCAA AGACTCTATC CTGACGCTAC GATTTCCTTA
30      601 TATCTAACAG CTTCCTGGTG TCGCAATGCT ATGGACAAAA AGAATCGGAA
651    GTTACTTAGT GATTGCGAAC TAAACCCCAA GATTGCTTGT TTGGACTTTA
701    ATCAGGGTGA TGTAATCAAA CAAGCAACTT GTGACTGTTG GATGGTGTAT
751    CATGGGGAGA ATGATCAAGG TACGTTGAAT CAGATTCAGG AAGAGTTAGA
801    AAAGTCAGGG GAGGAAACCC CTGGGATTCA TGTGGGGCAA AAGCCTCTTT
35      851 CACAATCCTT GTGGGATTTT TCTCCATTTT CATCTTTGGA GATGAAGGGA
901    GATAAAGAGA AAGCTCTAGA GTACTCTGAA TTAGAAAAAG AACAGCTATA
951    TTCTCGATTG GTATACGTAG GAGAGCGCTC TTCGTTCTT AGTTTGGGGT
1001    TTGGAGATAG TCGGTCAGGG ATCTTGATGG ACCCAAAACG GGTGCATGCT
1051    CCCTTATCTG AAGGGCATTG TTGTATTCC TACCTTGAG ACTTAGAAAA
40      1101 TCCCGGGTTA CAAAAACAA TTTTAGCGGC ATTTCTGAAT CCTAAGGAGT
1151    TGAGCAGTAC CATACTGCAA CCTATATCTC TAAATCTTAT CTAAATAGC
1201    AAAACTTACT TAAGGCAGCA CTTTGGCTTT TTTGAGAGGA TGAGCAGAAG
1251    TGATCGCAAT GTGGTTGTCT TGTATGTGA TTCTTGTTGG GGTACCGACT
1301    GGAAGGAGGA GCCAAGCTTC CAACACTTTA TTATGGAGCT AGAGTGTCTGA
45      1351 GGGTATTCTG ACTTCAATAT TTTTGCCTTT AGATCTAATA GCATGTGTGT
1401    AGAAGAACGT AGGATCTTAA ATGAAAGTTC TCAAGAGAAA GCCTTTACCA
1451    TGATTTCTCT TGAGGATTCA GTATCTCAAG GAGATATCCG CTGTTTGCAT
1501    TTGGCGTCTG AAGGAATGCT TTGTGGTAAA GAGTCTATG CTGTCGATGT
1551    CTATACGTCA GGATGCGCGA ACTTTATGAT GGAAGAAGTC TTAACCTTTG
50      1601 AGCGAGAAAT TAATCTGTGG AATAGAAAGC ATGCTCTTTG GAAAAGAGAA
1651    GTTAGAAAAC AGAAACAAGA AGCTGCTTTG GATCAAGACG AGAGCGAGAT
1701    TTACGTTTGT AATCAGCTGA CGGCGCAACA GAACCTCGCT TGTCTTTGA
```

The PSORT algorithm predicts inner membrane (0.2869).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 125A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 125B) and for FACS analysis.

These experiments show that cp6741 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 126

The following *C.pneumoniae* protein (PID 4376742) was expressed <SEQ ID 251; cp6742>:

```

5      1 LFVSNFIFV VMPPIYISSW ISTVRQHFKV AFDFSRPFCS RVTNPFALGVI
51     51 KAIPVGHIV MGMEWLVSSE VAGIITRSSF TSDVVQIVKT EKALGRDHIS
101    101 RVAEILQRR GTITPENQDK VHKGFPVCPF GRLKSEETLK LKPGEREGTL
151    151 DTVFSPIRTR VTRAYLQAPR PEIRTPISIVG SKLKTPQDFS QFVSLANETQ
10     201 RLHPEALVCL YLTGLNRESQ MCDTTAEKK QYLHNSGLDS RIQCKDSKED
251    251 DAGSPENPEL WIGYYSREQQ HNIDGQYIQQ CLGKSADPIP WIHVTEDTKD
301    301 FYYPPNFTSY SHTRQSTDPT SPPRLPESEG DKDSLYGQLS RSYHHEYMLG
351    351 LGLKPEDAGL LMDPDRIYAP LSQGHYCHSY LADIENEDLR TLVLSPFLDP
401    401 GNLSSDLRPF VAFNIARLPL ELDSLFFRLV AGQOEGRNIV TLAHGTPRPE
451    451 DLDPDSMNIL TRRLQMSGYS YLNIFSYSKR KMIVKERQFF GDRSEKGSFT
15     501 LILFEDPISA ADFRCLQLAA EGMVAKDLPS VADICASGCS CIQFSEMOSP
551    551 QAIEYRQWEA RVEDEAGEEA REPVIYSQDQ LSSMLTTOQN FVPSLDVAVK
601    601 QAIWRFRSKG LLTMRKALG EEFLTAIFY LGSQERNENM GKRTTEBHEV
651    651 VISFEELDRM VQVLPFAEVA DSGNDPTRPV PNPDSNPDS QNEGS*

```

The cp6742 nucleotide sequence <SEQ ID 252> is:

```

20      1 TTGTTTGTTF CTAATTTTAT TTTTTTGTGT GTTATGCCAA TTCCCTATAT
51     51 TTCTTCTTGG ATTCTACCG TTGACAGCA TTTTGTAAAG GCGTTTGATT
101    101 TCTCTCGTCC CTTTGTGTCT AGGGTTACGA ATTTGTCTTT AGGGGTCATC
151    151 AAGGCCATCC CTATTGTAGG ACATATTGTC ATGGGGATGG AGTGGTGTAGT
20     201 TTCTTCTGT GTTGCCGGGA TTATTACTAG GTCTCTCTTT ACCTCAGATG
25     251 TCGTTCAGAT TGTAAAGACT GAGAAGGCGT TAGGTCGAGA TCATATATCT
301    301 CGAGTGGCGG AGATATTGCA AAGAGAAAGG GGGACCATAA CTCCTGAGAA
351    351 TCAAGATAAG GTGCATGGGA AGTTTCTGT CTGTCTTTT GGTGCTTTAA
401    401 AATCCGAGGA AACTTTAAAA CTTAAGCCGG GAGAAAGAGA GGGAACTTTA
451    451 GATACTGTAT TTTCTCCGAT TCGCACGCGC GTGACTCGTG CGTACTTACA
30     501 GGCCCCCCGA CCCGAAATAC GTACGATTTT TATTGTGGGT TCGAAACTTA
551    551 AAACCTCTCA AGATTCTCG CAATTTGTGA GTCTCGCGAA TGAACCGCAG
601    601 AGACTGCATC CTGAAGCGTT AGTTTGTCTG TATTGTGACG GCTTGAATCG
651    651 CGAATCTCAG ATGTGCGGATA CAACTACTGC AGAGAAGAAG CAGTACCTAC
35     701 ATAACCTCAGG TCTCGACTCT AGAATCCAGT GCAAAGACAG TAAAGAAGAC
751    751 GACGCTGGCT CTCTGAAAA TCCCGAATT TGGATTGGCT ATTATTACG
801    801 AGAGCAACAG CATAATATAG ACGGGCAGTA TATTGACGAG TGTCTAGGGA
851    851 AGAGTGACAG TCCAATTCCT TGGATTTCATG TTACTGAAGA CACAAAGGAT
901    901 TTTTATTACC CACCAAACTT TACTTCATAC TCACATACAA GACAATCTAC
951    951 AGACCAACA TCGCCACCAA GACTCCCTGA AAGTGAGGGG GATAAGGATT
40     1001 CCTTGTACGG ACAACTGAGT CGATCGTATC ACCATGAGTA TATGCTTGGT
1051    1051 TTGGGATTAA AACCAGAGGA TGCAGGACTC CTGATGGACC CGGATAGAAT
1101    1101 CTATGCTCCT CTATCCCAAG GGCATTATTG TCATTCTTAC CTTGCGGATA
1151    1151 TAGAAAATGA GGATCTACGA ACTTTAGTCC TTTGCGCTTT CTTAGATCCT
1201    1201 GGCAATCTTA GTAGCGAGGA TCTTCGTCTT GTAGCATTC AATATCGCTAG
45     1251 ATTGCCATTA GAATTGGACT CGTTATTTT CCGCTTGTG GCGGGTCAGC
1301    1301 AAGAAGGGAG AAACATAGTT ACCCTTGCCC ACGGAACCTC TCGTCCAGAA
1351    1351 GATCTTGATC CTGACTCAAT GAACATCTG ACCAGAAGAT TACAAATGTC
1401    1401 TGGATATAGC TATTGGAACA TTTTCTCCTA TAAATCACGG AAAATGATTG
1451    1451 TAAAAGAACG TCAGTTCTTT GGAGATCGTT CTGAAGGGAA GTCTTTCACA
50     1501 TTGATCTTAT TTGAGGATCC CATTAGTGCA GCAGATTTCC GTTGTTTGCA
1551    1551 GCTAGCTGCA GAAGGTATGG TTGCTAAGGA TCTCCCAGC GTAGCAGATA
1601    1601 TTTGTGCTC TGGATGTTCC TGCATTCACT TTTCTGAGAT GCAGAGTCCT
1651    1651 CAGGCTATTG AATATAGACA ATGGGAGGCA CGTCTCGAAG ATGAAGCAGG
1701    1701 AGAAGAAGCC AGAGAACCAG TAATTTATTC TCAGGATCAA TTGAGCAGCA
55     1751 TGCTCACTAC ACAACAGAAT TTTGTATTTT CTCTAGATGC TGTGGTAAAA
1801    1801 CAGGCGATCT GGAGATTCCG TTCGAAAGGT CTCTTACTA TGGAAAGAAA
1851    1851 GGCAGTAGGC GAGGAGTTCT TAACTGCGAT ATTTCTCTAT TTAGGGAGTC
1901    1901 AGGAGCGTAA TGAGAATATG GGGAAAAGAA CTACCGAAGA ACATGAGGTC
1951    1951 GTTATCAGCT TCGAAGAGCT AGATCGCATG GTGCAAGTCC TCCAGCCGCA
60     2001 AGTCCCTGCA GATTCAGGCA ATGATCCTAC GCGTCCCGTT CCTAATCCAG
2051    2051 ATAGTAACCC TGATTCCTCG CAAAATGAAG GCAGTTAG

```

The PSORT algorithm predicts inner membrane (0.2338).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 126A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 126B) and for FACS analysis.

- 5 These experiments show that cp6742 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 127

The following *C.pneumoniae* protein (PID 4376744) was expressed <SEQ ID 253; cp6744>:

```

10      1 VIQHLLNFAL EETPSISVQY QBQEKLSPCD HSPEIGKKKR WNKLESFSTY
      51 CSLFMSVKDH YKLNLTGQNS LSGWLLDPYR VCAPLSSPYS CPSYLLDLQN
     101 KELRRSLST FLDPKNLTSE TFRSVSINFG NSSFGQRWSE FLRSVLHDEK
     151 EKHVAVVCND AKLLEEGLSP EALSLLLEEDL RESGYSYLN I LSVSPGVSK
     201 VQERQILRRD LQGRSFTVMI TDLPLGSEDI RSLQLASDRI LVSSSLDAAD
     251 ACASGCKVLV YENPNASWAQ ELENFYKQVE RRR*

```

- 15 The cp6744 nucleotide sequence <SEQ ID 254> is:

```

      1 GTGATACAAC ATCTTCTAAA CTTTGCTCTA GAAGAGACCC CTTCCATTTC
      51 CGTGCAATAC CAAGAACAAG AGAAGCTCTC TCCGTGCGAT CATTCCCCAG
     101 AAATAGGTAA AAAGAAAAGA TGAATAAGC TGAATCCTT CTCCACGTAT
     151 TGTCTCTGT TTAGTCTGT TAAGGATCAT TATAAGCTGA ATCTAGGAAT
     201 TCAGAAATCC CTGTCAGGT GGCTTCTGGA TCCCTATAGG GTTTGCGCGC
     251 CTTTATCTC ACCGTACTCG TGCTCTCCT ATCTTTTAGA TTTGCAAAAC
     301 AAAGAGCTAC GTCGTTCCCT TCTGTCAACG TTTCTAGACC CTAAAAATCT
     351 CACTAGCGAA ACATTCCGTT CTGTCTCTAT AAACCTTGGC AACTCTTCGT
     401 TTGGACAGAG ATGGTCAGAG TTTCTATCTC GTGTTCTGCA CGACGAGAAA
     451 GAAAAGCACG TAGCTGTTGT TTGTAATGAT GCAAACTTC TGGAGAAGG
     501 ATTGTCCCA GAGGCATTGT CTCTATTAGA AGAAGACTTA AGAGAATCAG
     551 GGTATTCGTA TCTAAACATT CTCTCGGTGA GCCCCGAAG AGTCTCCAAG
     601 GTTCAGGAAC GTCAGATTCT AAGGCGAGAT CTCCAAGGAC GGTCCCTTAC
     651 TGTCATGATT ACAGATCTTC CTTTAGGTAG CGAAGATATC CGTAGTTTAC
     701 AATTAGCCTC GGATAGGATT TTAGTCTCCA GTTCTCTTGA TGCCGCGGAT
     751 GCATGTGCTT CGGGATGTAA AGTCTTAGTC TACGAAAATC CAAATGCATC
     801 CTGGGCTCAG GAATTGAGA ACTTCTACAA ACAAGTTGAG AGAAGAAGGT
     851 AG

```

The PSORT algorithm predicts cytoplasm (0.3833).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 127A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 127B) and for FACS analysis.

These experiments show that cp6744 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

40 Example 128

The following *C.pneumoniae* protein (PID 4376745) was expressed <SEQ ID 255; cp6745>:

```

45      1 VACPSISWF TVVRQHFVNA FDFTHPVCSR ITNFALGIK AIPVLGHIVM
      51 GIEWLISWIP RHTVRHGMFT SDVSSAIKVE QTRGHNCLAP LEAYLSSLRV
     101 PISQEDLGKV HGRTPEDPFV DITPTEIVQL LPDEELSTVD EALQGVRSRL
     151 TYAYRSVEKP MIQDLALVGF GLRDSADLIN FVRLANGVQN HYPHTKVKLY
     201 LAKNLADVWD CEISEEEKGQ LRALGLDPKI ESISLTSAGL PSVPEVATVD
     251 FMITCYGKDQ EVQDP*

```

The cp6745 nucleotide sequence <SEQ ID 256> is:

```

1  GTGGCTTGTC CAAGTATTTT TCCTTGGTTT ACTGTCGTTT GACAGCATTT
51  TGTAAACGCC TTTGATTTCA CCCATCCCGT TTGTTCTCGG ATTACAAATT
101  TTGCTTTGGG GATCATTAAAG GCAATCCCGT TATTAGGACA CATTGTCATG
151  GGAATCGAGT GGTGATTTT CTGGATTCCC AGACACACCG TTGTCATGG
201  AATGTTTACT TCTGATGTCT CTAGTGCTAT TAAAGTAGAA CAAACACGGG
251  GTCATAATTG TTTAGCTCCC CTAGAAGCCT ATTTAAGTAG CTTGAGAGTC
301  CCCATTTCCT AAGAAGATCT AGGCAAAGTA CACGGGAGAA CCCCAGAAGA
351  TCCCTTCGTA GATATCACAC CCACAGAAAT TGTCCAACCT CTCCTGATG
401  AAGAACTCTC TACTGTAGAT GAGGCACTGC AAGGCGTTCG TAGTAGGTTA
451  ACCTATGCCT ATAGGTCCGT AGAGAAACCT ATGATTCAAG ATCTTGCTCT
501  TGTGGGTTTT GGTCTCCGAG ATTCTGCGGA CCTCATAAAT TTCGTGCGTC
551  TTGCTAATGG CGTGCAGAA CACTATCCCC ATACTAAAGT GAAGCTCTAT
601  TTAGCGAAGA ACTTGGCAGA TGTCTGGGAC TGTGAAATTT CTGAAGAGGA
651  AAAAGGGCAA CTCCGAGCTC TAGGTTTAGA CCCTAAAATA GAGAGTATAT
701  CCTTACGAG TGCAGGCTT CCTTCAGTGC CAGAAATCGC TACTGTGATG
751  TTTATGATTA CCGTTACGG GAAAGATCAG GAAGTCCAAG ATCCCTAG

```

The PSORT algorithm predicts inner membrane (0.2253).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 128A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 128B) and for FACS analysis.

These experiments show that cp6745 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 129

The following *C.pneumoniae* protein (PID 4376747) was expressed <SEQ ID 257; cp6747>:

```

1  MMKQGVGQDA KELYTFLSRG NEHYQPCLWF SLEELGLFLF DEKMLCAPLS
51  EDHYCHSYLV DLVDQHLKDL ILSMFLDPQN ISAGELLKVS INVGDSPSP
101  QQKDFLSMVL RDETGNVNVV VFKGVLSLPA TQVCKLVEEL NSKDYSYLNI
151  FSCHGDSSPQ LLFRKELEGT SGRYFTVICA LYLGDMDMS LQLASERIMV
201  SREFDLVDAY AARCKLLKID HTNWRPGTFS RHADFADAVD VSAGFNSREF
251  KLITQANQGI LESGELPLPS KTFWEGFLAF CDRVTVTRHF IPMLDAAIKQ
301  AVWTHKHPSL IDKECEALDL KTQCLPSIVS YLEYVTNSHE KTSKGPFIQK
351  EIIADCSPLK EALFPGSDED VPSTSEDPSP DHPSDLDS*

```

The cp6747 nucleotide sequence <SEQ ID 258> is:

```

35  1  ATGATGAAC AAGGAGTCGG GCAGGATGCT AAAGAGCTAT ACACATTTCT
51  ATCTCGTGGG AATGAGCATT ACCAACCGTG TCTATGGTTC AGTCTCGAAG
101  AGGAACTCGG ATTCCTTTTC GATGAAAAAA TGCTCTGCGC CCCTCTATCT
151  GAGGATCACT ATTGCCACTC GTATCTTGTA GATCTAGTGG ATCAACATTT
201  AAAGGATTTA ATATTATCGA TGTTTTGTGA TCCTCAGAAAT ATCTCAGCAG
251  GAGAACTCCT CAAGGTCTCT ATAAACGTTG GAGATTCTTT TTCTCCTCTA
301  CAACAGAAAG ATTCCTCTC GATGGTCTTA CGTGATGAAA CGGGAAAAAA
351  CGTCGTCGTG GTTTTAAAG GAGTTCCTCT CTACCCGCA ACCCAAGTCT
401  GCAAATTAGT AGAGGAATTG AACTCTAAGG ACTACTCCTA CCTCAATATA
451  TTTTCTTGTC ACGGAGATAG TAGTCCCTCAG CTTTATATCC GTAAGGAATT
501  AGAGGGAACT TCAGGGCGTT ATTTTACAGT GATTTGCGCT TTATATCTAG
551  GGGATACAGA CATGCGTAGT TTACAACCTG CTCTGAAAG GATCATGGTC
601  TCTAGAGAGT TTGATCTTGT AGATGCCTAT GCTGCAAGAT GCAAGCTCTT
651  GAAATCGAT CATACAAATT GGAGACCTGG AACTTTTCACT CGCCACGCCG
701  ATTTTCGAGA TGCTGTAGAC GTATCAGCAG GATTTAACTC AAGAGAATTT
751  AAAGTATTTA CGCAGGCGAA TCAAGGATC CTAGAGTCTG GAGAATCCCC
801  GCTCCCTTCA AAAACCTTCT GGAAGGATT CTTAGCATTC TGTGATCGAG
851  TGAATGTCAC GAGACACTTC ATTCCAATGT TAGACGCCG TATAAAGCAA
901  GCGGTATGGA CTCATAAACA TCCAGCTTGT ATAGATAAAG AGTGTGAAGC
951  CCTAGACTTG AAAACACAGT GCTTGCCATC TATCGTATCG TACCTTGAAT
1001 ATGTCACAAA CTCTACGAA AAAACATCGA AAGGCCGCTT CATAAAAAA
1051 GAGATTATCG CAGACTGTTC TCCTCTTAAA GAGGCGCTCT TCCCAGGTTT

```

1101 TGATGAAGAT GTTCCCTCTA CCTCTGAGGA TCCTTCAGAT GATCATCCTT
1151 CGGATCTTGA AGACTCTTAA

The PSORT algorithm predicts inner membrane (0.1447).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 129A) and also as
5 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 129B) and for FACS analysis.

These experiments show that cp6747 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 130

10 The following *C.pneumoniae* protein (PID 4376756) was expressed <SEQ ID 259; cp6756>:

1 MASGIGGSSG LGKIPKDNQ DRSRSPSPKG ELGSHEISLP PQEHGEEGAS
51 GSSHIHSSSS FLPEDQESQS SSSAASSPGF FSRVRSQVDR ALKSFGNFFS
101 AESTSQARET ROAFVRLSKT ITADERRDVD SSSAAATEAR VAEDASVSSE
151 NPSQGVPEVS SGPEPQRLFS LPSVKKQSG LRLVQTVRDR IVLPSSGAPPT
15 DSEPLSLYEL NLRSLSLRQE LSDIQSNDQL TPEEKAEATV TIQQLIQITE
201 FQCGYMEATQ SSVSLAEARF KGVETSDEIN SLCSLTDPE LQELMSDGDG
251 LQNLDEAD DLEAALSHTR LSFSLDDNPT PIDNNPTLIS QEEPIYEEIG
301 GAADPQRTRE NWSTRLWNQI REALVSLGGM ILSILGSILH RLRIARHAAA
351 EAVGRCCCTCR GEECTSSEED SMSVSGSPSEI DETERTGSPH DVPRRNGSPR
401 EDSPLMNALV GWAHKGAKT KESSESSTPE ISISAPIVRG WSQDSSVSFI
451 VMEDDHIFVD VPRKKGDIYD VPSSPRWSPA RELEDVFGD YEVPIITSAEP
501 SKDKNIYMTF RLATPAIYDL PSRPGSSGSS RSPSSDRVRS SSPNRRGVPL
551 PPVPSAMSE EGSYEDMSG ASGAGESDYE DMSRSPSPRG DLDEPIYANT
601 PEDNPFQTQRN IDRIQLQERSG GASASPEVEPI YDEIPWIHGR PPATLPRPEN
651 TLTNVSLRVS PGFGPEVRAA LLESVSASVM VEAEISVPPT EPGDGESEYL
701 EPLGGLVATT KILLQKGWPR GESNA*
751

The cp6756 nucleotide sequence <SEQ ID 260> is:

1 ATGGCATCAG GAATCGGAGG ATCTAGTGGA TTAGGAAAAGA TTCCACCTAA
51 AGATAATGGG GATAGAAGTC GATCGCCCTC TCCTAAGGGA GAACCTGGCA
101 GCCACGAGAT TTCCCTGCCT CCTCAAGAAC ATGGAGAGGA AGGAGCTTCA
151 GGATCTTCGC ATATACATAG CAGTTCCTCT TTTCTACCAG AAGATCAGGA
201 GTCTCAGAGC TCTTCTTCGG CAGCTTCTAG CCCGGGATTT TTTTCTCGCG
251 TACGTTCTGG GGTAGACAGG GCCTTAAAT CATTTGGCAA CTTTTTTTCC
301 GCAGAGTCTA CGAGTCAAGC GCGTGAAACG CGACAAGCTT TTGTTAGATT
351 ATCAAAAACC ATCACCAGCGG ATGAGAGACG GGATGTGAT TCATCAAGTG
35 CTGCTGCTAC AGAAGCCCGA GTGGCAGAGG ACGCGAGTGT TTCAGCGCAA
401 AATCCTTCTC AGGGGGTTCC AGAAACCTCT TCTGGACCAG AACCTCAGCG
451 TTTATTTTCT CTTCCTTCAG TAAAAAACA GAGCGGTTTG GGTGCGTTGG
501 TACAGACAGT TCGCGATCGC ATAGTACTTC CTAGTGGGGC TCCACCTACA
40 GACAGCGAGC CTTTAAGTCT CTACGAGCTA AACCTCCGTT TGAGTAGTTT
601 ACGTCAGGAG CTCTCTGACA TACAAAGTAA TGATCAGTTG ACTCCAGAGG
701 AAAAAGCAGA AGCCACAGTT ACCATACAAC AGCTGATCCA AATTACAGAA
751 TTCCAATGCG GCTATATGGA GGCAACACAA TCTTCGTAT CTCTAGCAGA
801 AGCTCGTTTT AAGGGGGTAG AAACCTAGTGA TGAGATCAAT TCCCTCTGTT
45 CAGAACTGAC AGATCCTGAG CTTCAAGAAC TCATGAGTGA TGGAGACTCT
901 CTTCAAAACC TATTAGATGA GACTGCCGAC GATTTAGAAG CTGCTTTGTC
951 CCATACTCGA TTGAGTTTTT CTTTAGACGA TAATCCAAC CCGATAGACA
1001 ATAATCCAAC TCTGATTTCT CAAGAAGAGC CTATTATGA GGAAATCGGA
1051 GGAGCTGCAG ATCCTCAAAG AACTCGGGAA AACTGGTCTA CAAGATTATG
50 GAATCAGATT CGCGAGGCTC TGGTTTCTCT TTTAGGAATG ATTTTAAGCA
1151 TTCTAGGGTC CATCTGTCAC AGGTGCGTA TTGCTCGTCA TGCAGCTGCT
1201 GAAGCAGTGG GTCGTTGTTG CACGTGCCGA GGAGAAGAGT GTACTTCTTC
1251 TGAAGAGGAC TCGATGTCGG TGGGGTCTCC TTCAGAAATT GATGAACTG
1301 AAAGAACGGG CTCCTCCGAT GACGTTCCAC GCAGAAATGG AAGTCCACGT
55 GAAGATTCTC CATTGATGAA TGCTTAGTA GGATGGGCAC ATAAGCACGG
1401 TGCTAAAACC AAGGAGAGTT CAGAATCAAG TACCCCGAA ATTTGATTT
1451 CTGCTCCCAT AGTGAGAGGT TGGAGTCAAG ACAGTTCCGT CAGTTTTATT

5
 10
 15

```

1501 GTTATGGAAG ATGATCATAT TTTCTATGAT GTTCCTCGTA GAAAAGATGG
1551 AATCTATGAC GTTCCTAGTT CCCCTAGATG GAGTCCTGCG CGAGAGTTGG
1601 AAGAGGATGT TTTTGGAGAT TATGAAGTTC CTATAACCTC TGCTGAACCA
1651 TCTAAAGACA AGAACATCTA CATGACACCT AGATTAGCAA CTCCTGCTAT
1701 CTATGATCTT CCTTCACGTC CAGGATCGTC TGGGAAGCTCA CGTTCTCCGT
1751 CTTCAGATCG CGTACGAAGC AGCTCACCAA ATAGACGGGG TGTGCTCTCT
1801 CCTCCAGTTC CTTACACCTGC TATGAGTGAG GAGGGGAGCA TTTATGAGGA
1851 TATGAGCGGT GCTTCAGGTG CAGGTGAAAG TGATTATGAA GATATGAGCC
1901 GTTCCCCCTC TCCTAGAGGC GACTTGGATG AAGCCATATA TGCTAATACT
1951 CCTGAAGATA ATCCATTAC TCAGAGAAAT ATAGATAGAA TTTTACAGGA
2001 GAGGTCAGGC GGTGCTTCCG CTTCTCCTGT AGAGCCTATT TATGATGAGA
2051 TCCCATGGAT TCATGGCAGG CCCCCTGCTA CACTTCCAAG ACCCGAGAAT
2101 ACATTGACTA ATGTTTCGCT TAGAGTGAGC CCAGGGTTTG GACCAGAAAT
2151 AAGAGCCGCT TTGCTTAGCG AGAGCGTGAG TGCTGTTATG GTCGAAGCAG
2201 AGAGTATTGT TCCTCCAACA GAGCCGGGGG ACGGAGAATC AGAATATCTA
2251 GAGCCCTTAG GGGGACTTGT AGCTACAACG AAAATCTTAC TACAAAAGG
2301 ATGCCCTCGT GGAGAGTCGA ATGCTTAG
  
```

The PSORT algorithm predicts inner membrane (0.3994).

20

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 130A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 130B) and for FACS analysis.

These experiments show that cp6756 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 131

25

The following *C.pneumoniae* protein (PID 4376761) was expressed <SEQ ID 261; cp6761>:

30

```

1 MTVAEVKGTf KLVCLGCRVN QYEVQAYRDQ LTILGYQEV L DSEIPADLCI
51 INTCAVTASA ESSGRHAVRQ LCRQNPTAHI VVTGCLGESD KEFFASLDRQ
101 CTLVSNKEKS RLIEKIFS YD TTFPEFKIHS FEGKSRAFIK VQDGCNSFCS
151 YCIIPLYLRGR SVSRPAEKIL AELAGVVDQ YREVVIAGIN VGDYCDGERS
201 LASLIEQVDR IPGIERIRIS SIDPDDITED LHRAITSSRH TCPSSHVLVQ
251 SGSNSILKRM NRKYSRGDFL DCVEKFRASD PRYAFTTDVI VGFPGESDQD
301 FEDTLRIED VGFIVKHSFP FSARRRTKAY TFDNQIPNV IYERKKYLA E
351 VAKRVGQKEM MKRLGETTEV LVEKVTGQVA TGHSPYFEKV SFPVVGTVAI
401 NTLVSVRLDR VEEGLIGE I V*
  
```

35

The cp6761 nucleotide sequence <SEQ ID 262> is:

40
 45
 50
 55

```

1 ATGACGGTTG CGGAAGTCAA AGGAACATTT AAGCTGGTCT GTTTAGGCTG
51 TCGGGTGAAT CAGTATGAGG TCCAAGCATA TCGCGACCAG TTGACTATCT
101 TAGGTTACCA AGAGGTCCCTG GATTCTGAAA TCCCTGCAGA TTTATGCATA
151 ATCAATACGT GTGCTGTCAC AGCTTCTGCT GAGAGTTCCG GTCGTCATGC
201 TGTGCGTCAG TTATGTCTGTC AGAACCCTAC AGCACATAAT GTTGTCACAG
251 GTTGTGTTGGG GGAATCTGAC AAAGAGTTTT TTGCTTCTTT GGATCGGCCA
301 TGCACACTTG TTTCCAATAA AGAAAAATCC CGACTTATAG AAAAAATTTT
351 TTCCTATGAT ACGACCTTCC CTGAGTTCAA GATCCATAGT TTTGAGGGAA
401 AGTCTCGAGC TTTTATTAAA GTTCAAGATG GCTGTAATTC TTTTGTCTCG
451 TACTGCATTA TTCCTTATTT GCGGGGGCGT TCGGTTTCTC GTCCTGCTGA
501 GAAGATTTTA GCTGAAATCG CAGGGGTGTG AGACCAAGGA TATCGCGAAG
551 TTGTAATTGC AGGAATTAAT GTTGGAGATT ATTGCGATGG AGAGCGTTCA
601 TTAGCCTCTT TGATTGAACA GGTGGACCGG ATTCTTGGA TTAGAGGAT
651 TCGAATTTCC TCTATAGATC CTGATGATAT CACTGAAGAT CTGCACCGTG
701 CCATCACCTC ATCGCGTCAC ACTTGTCTCT CGTCACACCT GTTCTCTCAA
751 TCGGGGTCGA ATTCATTTT AAAGAGAATG AACCGGAAGT ATTCTCGCGG
801 AGATTTTTTA GATTGTGTAG AGAAGTTCCG TGCTTCTGAT CCTCGCTATG
851 CCTTTACTAC AGATGTGATT GTCGGATTTC CTGGAGAGAG TGATCAAGAT
901 TTTGAAGATA CTTTGAGAA TATTGAAGAT GTAGGCTTTA TTAAGTGCA
951 TAGTTTCCCT TTCAGTGCTC GTCGTCGTAC TAAGGCATAT ACTTTTGATA
1001 ATCAGATTCC CAATCAGGTG ATCTATGAGA GGAAGAAGTA TCTTGCTGAG
1051 GTTGCTAAGA GGTAGGCCA GAAAGAGATG ATGAAGCGTT TAGGAGAGAC
  
```


1101 TACAGAGGTG CTTGTTGAGA AAGTAACGGG GCAGGTTGCT ACGGGTCACT
 1151 CTCCTTATTT TGAAGAGGTT TCTTCCCTG TTGTAGGAAC GGTAGCTATC
 1201 AACACTCTAG TTTCTGTGCG TCTTGATAGG GTAGAGGAAG AAGGGCTGAT
 1251 TGGGGAGATT GTATGA

5 The PSORT algorithm predicts inner membrane (0.1574).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 131A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 131B) and for FACS analysis.

10 These experiments show that cp6761 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 132

The following *C.pneumoniae* protein (PID 4376766) was expressed <SEQ ID 263; cp6766>:

1 MATSVFVTSS TSVGEANSSN ERFTESTRM YYAALVLGAL SCLIFIAMIV
 51 IFPOVGLWAV VLGFGALGCLL LSLAIVFAVS GLVLGKTLEP SREATPPEIV
 101 AQKEWTQQD VLGNEYWRSE LISLFLRGDL HESLIVDSKD RSLDIDQSLQ
 151 NILKLEPLST TSLSLKKDCV HINIILHLVR QWNLLGVDLS PEVTAHABEL
 201 LLFLIEEQYY SPDILKLIRY GDALQATSPL MDWADSGSFS VDADGVFSCR
 251 REECSPEDAL AQFDLLALE NPDRRLKDS FLTYIWSSSF FEKFLHRHLE
 301 SLQRKLPETA IDVARYEAI QTFLSRYFQK LDLINAMSLD WGYNCAEGEK
 351 CYESANQRLD NLFIAFSSSV PAMKRLFDKY GSVVRVDRRQ IREQILSNT
 401 ILENESGFLC SLYEYPLSYL IDWAVLLDCV RGTEISLEDQ ADYTVCLQGL
 451 DSMLSQFASR LQSGQKVLNP RDVLSEQAAV MLVHGLAAQG VSFQGLKALM
 501 YLTAVPQRMW LGALPLFESF PVFNRMKFL GESLGD*

The cp6766 nucleotide sequence <SEQ ID 264> is:

25 1 ATGGCAACCT CTGTTCTGT AACTTCATCT ACTTCTGTAG GAGAGGCTAA
 51 CTCCTCCAAC GAAAGATTTA CTGAACGAAC ATCGCGAATG TATATACGAG
 101 CTTTAGTCCT AGGGGCTTTG AGCTGTTTAA TTTTATTGTC TATGATTGTC
 151 ATTTCCTCAC AGGTCGGATT GTGGGCTGTG GTCTCGGGT TTGCTCTTGG
 201 ATGTTTACTT TTAAGCTTAG CTATCGTTT TGCTGTCTCC GGTCTCGTTT
 251 TAGGCAAGAC TTAGAACCT AGTCGAGAAG CGACTCCTCC AGAAATTGTT
 301 GCGCAAAAGG AGTGGACTAC ACAACAAGAT GTCTTAGGGA ATGAGTATTG
 351 GCGTTCCGAG TTGATTTCTT TGTCTTACG AGGGGATCTC CACGAATCTC
 401 TGATTGTTGA TTCTAAGGAT CGATCTTTAG ATATTGATCA GAGTTTACAA
 451 AATATATGTA AACTTGAGCC CCTATCTACG ACACTTTCGC TGTAAAGAA
 501 AGATTGTGTC CACATCAATA TCATTTTACA TTTAGTGAGA CAGTGGAAC
 551 TACTGGGAGT GGATCTTAGT CCTGAAGTCA CTGCGCACGC CGAGGAACTT
 601 CTACTCTTTT TGATAGAAGA GCAGTATTAC TCTCTGATA TTTTGAAATT
 651 GATTCGCTAC GGAGATGCTT TACAAGCAAC GTCTCCTTTG ATGGATTGGG
 701 CAGATTCAGG TTCCTTTAGT GTAGACGAG ACGGGGTATT TAGCTGTGCG
 751 AGAGAAGAAT GTTCTCCTGA GGATGCTTTG GCGCAATTCT ATCTTCTTTT
 801 GCGCTTGGAA AATCCCGACA GACGCTTCTT AAAGGATTCT TTTCTTACCT
 851 ACATTGCTG GTCTTCATTT TTTGAGAAGT TTTTACATCG CCATCTAGAG
 901 AGCTTGCAAA GAAAGCTCCC AGAGACAGCG ATCGATGTCG CCCGCTATGA
 951 AGCACAAATA CAAACATTTT TCTCTCGCTA TTTTCAGAAG CTCGATTGTA
 1001 TAAACGCAAT GTCCTTAGAT TGGGGATATA ACTGTGCTGA GGGAGAAAAA
 1051 TGTTATGAGA GCGCAAATCA AAGATTAGAC AACCTATTTA TTGCTTTTTC
 1101 TTCTTCTGTT CCTGCTATGA AGCGGCTCTT TGACAAATAT GGTTCGTGTT
 1151 TACGGGTAGA TCGTAGGCAG ATTCGTGAGC AGATCTTTTC GAACACTGAA
 1201 ATCTTAGAAA ATGAGTCAGG GTTCTCTGTC AGTTTGTATG AATATCCTTT
 1251 ATCTTATTG ATAGATTGGG CTGTTTGTCT AGACTGTGTT CGCGGTACCG
 1301 AAATCTCTCT AGAAGATCAG GCCGATTACA CCGTTTGTCT GCAAGGCTTG
 1351 GATCTATGTT TATCTCAATT TCGGAGTCGT TTACAGTCTG GACAAAAAGT
 1401 ATTGAATCCT AGAGATGTTT TAAGTGAACA GGCTGCGGTT ATGCTTGTTC
 1451 ATGGCTTGGC AGCACAGGGC GTGTCGTTTC AAGGATTGAA AGCTTTGATG
 1501 TATTGACAG CCGTTCCCCA AAGAATGTGG TTAGGAGCAT TGCCCTTTAT
 1551 TGAATCTTTT CCTGTCTTTA ATCGGATGAA AGAATTTCTT GGGGAATCTC
 1601 TGGGAGACTA G

The PSORT algorithm predicts inner membrane (0.6158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 132A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 132B) and for FACS analysis.

- 5 These experiments show that cp6766 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 133

The following *C.pneumoniae* protein (PID 4376804) was expressed <SEQ ID 265; cp6804>:

10 1 MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLLIEGK
 51 LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
 101 ATLESRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLQOTP ENYDGLLLIG
 151 DAALQHPVLP GFVTDLASG WYDLTKLPFV FALLHSTSW KEHPLPNLAM
 201 EEALQQFESS PEEVLKEAHQ HTGLPPSLQ EYYALCQYRL GEEHYESFEK
 251 PREYYGTLYQ QARL

- 15 The cp6804 nucleotide sequence <SEQ ID 266> is:

 1 ATGTCTAACC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
 51 TAATTCCTTT CCGCTGTCCC TACAACATCAT AAAAAGAAAC GATATTGCGT
 101 GTGTTCTTGC TCCCCCTGCA GACCTCCTCA ACTTGCTAAT CGAAGGGAAA
 151 CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTGGG
 201 GTATGTCCCC GGCTTTGGAA TTGCAGCAAA CCAACGTATC CTCAGTGTAA
 251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC
 301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAAG TGCTTTGTCG
 351 TCATCTCTGG CGCATCCCAA CTCCTCATAT CCTAAGATTG ATAACACAA
 401 AAGTACTCAG ACAAACCCCT GAAATTTATG ATGGCCTCCT CCTAATCGGA
 451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTGTGTA CCTATGACCT
 501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTTGCTCTTC
 551 TTCTACACAG CACCTCTTGG AAAGAATC CCAACCCAA CCTTGCGATG
 601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA
 651 AGCTCATCAA CATACAGGTC TGCCCCCTTC TCTTCTTCAA GAATACTATG
 701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAA
 751 TTCCGGGAAT ATTATGGAAC CCTCTACCAA CAAGCCCGAC TGTA

The PSORT algorithm predicts inner membrane (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 133A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 133B) and for FACS analysis.

35 These experiments show that cp6804 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 134

The following *C.pneumoniae* protein (PID 4376805) was expressed <SEQ ID 267; cp6805>:

40 1 MSSLLSCGRI EPTRVTCSLK TYLEDTSQNG LSTRLVASV IFLCALLIIL
 51 VCVALSSLIP SIMALATSFT VMGLILFVMS LLGDVAISY LTYSTVTSYR
 101 QNKRAFEIHK PARSVYEGV RHWDLGRSSL GTGEIPIVRT LFSFFQNHGL
 151 NHALAAKIFL FMEHFSPEPP NEPLVDWACL IRDFRPHVSS LCFVIEKQGS
 201 SLRTKEGNTI CEAFRSYDA HFAMVDCYRL IHSKLIIEKM GLKNIDIIPS
 251 VMVREDYPSR PGEGYREGLL RMYGGKAL*

The cp6805 nucleotide sequence <SEQ ID 268> is:

5
10
15

```

1  ATGTCATCAC  TACTGAGCTG  CGGAAGAATA  GAGCCGACTC  GGGTTACCTG
51 TAGCTTAAAG  ACGTATCTTG  AGGATACGAG  TCAGAATCAG  TTGAGCACAC
101 GTCTAGTTCG  GGCAAGTGTC  ATCTTTTAT  GCGCATTGTT  GATCATTTTG
151 GTTTGTGTGG  CCTCTCTAG  TTTGATTCCA  AGCATTATGG  CCTTGGCGAC
201 CTCTTTTACG  GTAATGGGGT  TAATCTTTT  TGTGATGTCA  CTTCTTGGTG
251 ACGTTGCAAT  TATAAGTTAT  CTTACTTATA  GCACTGTTAC  GAGTTACCGG
301 CAAAATAAGA  GAGCTTTTGA  GATTCACAAG  CCCGCTCGCT  CCGTTTACTA
351 CGAGGGGGTC  CGCCATTGGG  ATTTAGGACG  ATCATCTTTA  GGCACAGGCG
401 AGATTCCCTAT  AGTAAGGACG  TTATTCTCTC  CATTTAGAA  CCATGGTCTT
451 AACCATGCCT  TAGCTGCTAA  AATTTTCCTA  TTTATGGAGC  ATTCAGGCC
501 TGAGCCACCG  AACGAGCCTT  TGGTGGATTG  GGCCTGTTTG  ATTCGGGATT
551 TTAGGCCTCA  CGTCAGTTCT  TTGTGCTTTG  TTATTGAAAA  ACAAGGGTCA
601 TCGCTGAGGA  CTAAGGAAGG  CAATACGATT  TGTGAGGCTT  TCCGCTCTGA
651 TTACGACGCC  CATTTTGCTA  TGGTAGATTG  CTACCGGTTG  ATCCACTCTA
701 AGTTGATTAT  AGAGAAAATG  GGATTGAAGA  ATATCGATAT  CATTCGAGT
751 GTCATGGTTC  GTGAAGATTA  TCCTAGCCGT  CCTGGGGAGG  GCTATCGCGA
801 AGGCCTATTA  CGTATGTATG  GTGGCAAGGG  GGCTCTGTGA

```

The PSORT algorithm predicts inner membrane (0.711).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 134A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 134B) and for FACS analysis.

These experiments show that cp6805 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 135

25 The following *C.pneumoniae* protein (PID 4376813) was expressed <SEQ ID 269; cp6813>:

30

```

1  MSGPSRTESS  QVSVLSYVPR  DKEIAPKKQF  TIAKISTLAI  LASLALGALV
51 AGISLTIVLG  NPVFLALLIT  TALFSVVTF  VYHQMTSKVS  SNWQKVLEQN
101 FKPLGKAWQE  KNVDCYSNEM  QFYNNHLNPK  FKVAIQTDAS  QPFQPTFLTG
151 LRVIEKNQST  GIIFNPVGPT  NLIDNTATNL  STILYSTLKD  KSVWDTCKQR
201 EGGPAKGEDP  FSPTEVRVVK  LPNEALDQTF  NLNLSSAEKK  SILPTFLGHV
251 CGPKSEELPN  QQEYYRQALL  AYENCLKAAI  ESHAIVLALP  LFTSVYEVPP
301 EEILPKEGTF  YWDNQTOAFC  KRALLDAIQN  TALRYPQRSL  LVILQDPFNT
351 IESQSRSEB*

```

The cp6813 nucleotide sequence <SEQ ID 270> is:

35
40
45
50
55

```

1  ATGTCAGGAC  CCTCACGTAC  TGAGAGCTCT  CAAGTTTCTG  TACTATCCTA
51 TGTGCCTCGG  GATAAAGAAA  TTGCTCCTAA  AAAACAGTTT  ACCATAGCAA
101 AAATATCCAC  TCTTGCAATC  CTAGCTTCTT  TAGCTTTAGG  AGCTTTGGTG
151 GCTGGAATCT  CTTTAACGAT  AGTATTAGGG  AACCCTGTAT  TTTTGGCTCT
201 TCTCATTACC  ACGGCCCTCT  TCTCAGTTGT  AACCTTCTTA  GTCTACCACC
251 AAATGACCTC  AAAGGTATCT  TCTAACTGGC  AGAAAGTTCT  AGAGCAAAAC
301 TTCAAGCCTT  TGGGAAAAGC  GTGGCAAGAA  AAAAACGTAG  ACTGCTACTC
351 AAACGAGATG  CAATTTTACA  ATAATCACCT  GAACCCTAAG  TTCAAGGTAG
401 CGATACAAAC  AGATGCGTCT  CAACCATTTC  AGCCTACTTT  CTTAACTGGA
451 CTTAGAGTGA  TCGAAAAAAA  TCAATCCACA  GGGATCATCT  TTAATCCCGT
501 AGGCCCAACG  AATCTGATCG  ACAACACTGC  AACGAACCTC  TCTACTATCC
551 TTTACTCCAC  CCTAAAAGAT  AAAAGCGTGT  GGGATACATG  CAAGCAACGC
601 GAAGGGGGTC  CCGCAAAAGG  AGAAGACCCC  TTTTCCCCCTA  CCGAAGTGAG
651 AGTAGTAAAA  CTTCCAACAG  AAGCTCTAGA  TCAAACGTTT  AATCTAAATT
701 TAAGCTCTGC  AGAAAAGAAA  AGTATTCTTC  CGACCTTTTT  AGGCCACGTA
751 TGCGGCCCTA  AATCTGAAGA  GTTACCAAAT  CAGCAAGAAT  ATTATCGCCA
801 AGCTTTACTA  GCGTACGAGA  ACTGCCTTAA  AGCAGCTATA  GAAAGTCATG
851 CAGCAATCGT  TGCTCTTCCT  CTCTTACTT  CGGTCTATGA  AGTGCCCTCA
901 GAAGAGATTG  TTCCTAAAGA  AGGCACTTTC  TATTGGGACA  ACCAAACTCA
951 AGCGTTTTCG  AAACGCGCTT  TATTGGACGC  TATTCAAAT  ACGGCCCTAC
1001 GCTATCTTCA  AAGATCTTTA  CTTGTTATAC  TCCAAGATCC  TTTTAATACT
1051 ATAGAATCAC  AAAGTCGTTT  TGAGGAGTAA

```

The PSORT algorithm predicts inner membrane (0.4291).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 135A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 135B) and for FACS analysis.

- 5 These experiments show that cp6813 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 136

The following *C.pneumoniae* protein (PID 4376844) was expressed <SEQ ID 271; cp6844>:

```

10      1 MWRVVLRLFI IFILGRAVFP LRASESFSWE TSTCLTVLGI PFIDIILTTN
      51 EDFVAQCGLQ IGTISSSTNNA KIKEIFLIYK EKFPASISF KRKEPLNLSQ
      101 SHLSDLGILC MRNGETYABG MANKENGPAI KQPKDLRLVL RCPNQPDLLL
      151 YSEKEAEKGI ETNTCLCNQG YTLDDGQLIL YGDSIEKFLK ETKRKNNHTL
      201 VDLCDSDQVVT TFLGRFWSLL NYVQVFLSE DSAKILAGIP DLAQATQLLS
      251 HTVPLLFYIT NDSIHIEQG KESSFTYNQD LTPILGLFLF GYINRGSMSEY
15      301 CFNCAQSSLG ET*
```

The cp6844 nucleotide sequence <SEQ ID 272> is:

```

      1 ATGTGGCGCG TTGTCCTCAG ATTCCTTATA ATTTTATCT TGGGAAGAGC
      51 CGTCTTCCCT CTAAGAGCTT CAGAAAGCTT CTCCTGGGAA ACATCGACCT
      101 GTTTAACAGT GCTAGGGATT CCTTTCATAG ATATTATCCT CACAACGAAT
      151 GAGGACTTGT TTGCCAGTGT CGGCCTGCAA ATAGGAACCA TTTCTTCGAC
      201 TAATAACGCA AAAATAAAAG AAATTTTATG GATATATAAG GAAAAATTTT
      251 CAGAAGCCTC TATCAGTTTC AAACGAAAAG AACCTCTAAA CCTTTCCCAA
      301 TCCCATCTCT CCGATTTAGG TATTTTATGT ATGCGTAACG GAGAAACTTA
      351 CGCTGAGGGA ATGGCAAATA AAGAAAACGG ACCCGCTCTA AAACAACCCA
      401 AGGATCTAAG ATTAGTTTTA CGTTGTCCTA ACCAACCAGA TACCTTGCTC
      451 TACTCGGAAA AAGAAGCAGA AAAGGGCATA GAAACAAATA CTTGCCTATG
      501 CAATCAGGGA TACACACTCC TGGATGGGCA ATTGATTCTC TACGGGGATA
      551 GTATAGAAAA GTTTCGAAA GAGACCAAAA GAAAGAATAA CCACACGCTT
      601 GTTGATCTTT GTGACTCACA AGTCGTGACC ACGTTCCTCG GTCGCTTTTG
      651 GTCTCTTCTA AACTACGTTT AAGTTCCTTT CCTATCTGAA GACTCCGCTA
      701 AAATTTCTTG GGGCATCCCA GACCTAGCTC AAGCTACGCA ATTGCTTTCC
      751 CACACCGTAC CTTTGCTTTT TATTTATACC AACGATTTCT TTCACATCAT
      801 AGAACAAAGC AAAGAAAGTA GTTTTACCTA TAACCAAGAT TTAACAGAGC
      851 CCATTTTAGG ATTTCTCTTT GGTACATAA ATCGCGGCTC TATGGAATAC
35      901 TGCTTTAATT GTGCACAGTC TTCATTAGGA GAAACCTAA
```

The PSORT algorithm predicts inner membrane (0.1786).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 136A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 136B) and for FACS analysis.

- 40 These experiments show that cp6844 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 137

The following *C.pneumoniae* protein (PID 4377201) was expressed <SEQ ID 273; cp7201>:

```

45      1 VLVGICPSLY PEHPRSFFYYR VSGDIGSRFD DRGFVNSGVE TLPYSSGSFG
      51 IFWISFTDPT FNFAIVNTFM RTAGINEVSR PMTQDTETSL IEMRDLSEQQ
      101 EANNPDSLEQ EESLMGIIVGH TVGGVSMTVT SSPNIFYRIQ TLLGLPETLA
      151 EAENPTFPN STIDSLAEIM MNLVRISDAV SIFWIFPIVD TTYNGVLLAV
```

201 CIGFFGINGI CSTFLMLTNP RSRRDRWRNL RIMVLCYRSL GSGMNLFDLS
 251 NNVRMAARRH VTSCTVALYA MVTLFGWTV IQDALQYGFSP SVRDAFYRYC
 301 LRHRYCLTQR NEDSLQTTGT RFQVTRTHLE DQOMVASILN LSVFGLFFFGF
 351 VGLMTTFGGL EISPSRWDA ANNRTVGIF*

5 The cp7201 nucleotide sequence <SEQ ID 274> is:

1 GTGCTCGTTG GTATCTGTCC TTCTCTATAT CCAGAACATC CTCGCTCCTT
 51 TTATTATCGT GTTTCCTGGAG ATATAGGCTC CCGATTTCGAC GATAGAGGAT
 101 TTGTAAACTC TGGAGTCGAA ACCCTGCCAT ACTCTTCAGG CAGCTTTGGG
 151 ATTTTGTGGA TCTCGTTTAC GGATCCACACA TTTAATTTTG CTATCGTAAA
 201 TACCTTTATG CGAACTGCAG GGATCAATGA AGTCTCTAGA CCCATGACAC
 251 AAGATACAGA AACTTCATTG ATAGAAATGA GAGACCTAAG TGAACAACAA
 301 GAAGCGAATA ACACAGATTC TTTAGAGCAA GAAGAGAGCT TAATGGGTAT
 351 TGTAGGACAT ACTGTGGGAG GAGTTTCCAT GACCGTGACC TCCAGTCCAA
 401 ATATCTTTTA TCGTATACAA ACACCTCTGG GACTGCCAGA GACTCTTGCA
 451 GAAGCTGAAG AAAATCCTAC CTTCCTCAAAT TCTACTATAG ATAGCCTTGC
 501 AGAAATAATG ATGAACCTCG TAAGGATCTC TGATGCTGTC TCTATTTTCT
 551 GGATTTTTC TATCGTAGAT ACTACATATA ATGGAGTTT ATTAGCCGTC
 601 TGTATCGGCT TCTTCGGAAT CAATGGGATT TGTTCACCGT TCCTTATGCT
 651 TACGAATCCA CGCTCTCGTC GAGATAGATG GAGGAATTTA CGCATCATGG
 701 TTCTTTGCTA TCGTCTTTTG GGAAGCGGAA TGAATCTCTT TGATCTTAGC
 751 AATAATGTGC GCATGCCAGC ACGTAGGCAT GTGACATCAT GTACAGTAGC
 801 TCTCTATGCT ATGGTCACTC TATTGGATG GACAGTAGCA ATACAAGATG
 851 CTTTGCAATA TGGTTTCCCT AGCGTTCCGG ATGCCTTCTA TAGATATTGC
 901 TTACGCCACA GATATTGCTT AACTCAAAGA AACGAAGACT CTCTGCAAC
 951 TACAGGAACG CGCTTTCAGG TTACCCGTAC ACATCTAGAA GATCAACAGA
 1001 TGGTGGCTTC TATTTTGAAT TTGAGTGTTT TTGGGCTCTT TTTTGGATT
 1051 GTAGGGCTAA TGACCACGTT TGGAGGATTA GAAATCTCAC CATCTTGTCG
 1101 GTGGGATGCA GCAATAAACC GAACGGTAGG TATTTTTTAG

The PSORT algorithm predicts inner membrane (0.3102).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 137A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 137B) and for FACS analysis.

These experiments show that cp7201 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 138

The following *C.pneumoniae* protein (PID 4377251) was expressed <SEQ ID 275; cp7251>:

1 MAPIHGSNAF VEDILHSHPS PQATYFSSTR AQKLHEFKDR HPVLTRIASV
 51 IIKIFKVLIG LIILPLGIYW LCQTLCTNSI LPSKNLLKIF KKQPNKTLK
 101 TNYLHALQDY SSKNRVSMR RVPILQDNVL IDTLEICLSQ APTNRWMLIS
 140 151 LGSDCSLEEI ACKEIFDSWQ RFAKLIGANI LVYNYPGVMS STGSSSLKDL
 201 ASAHNICTRY LKDKQGPFGA KEIITYGYSL GGLIQAEALR DQKIVANDDT
 251 TWIAVKDRCP LFISPEGFHS CRRIGKLVAR LFGWGTKAVE RSQDLPLEI
 301 FLYPTDSLRR STVRQNKLLA PELTLAHAIK NSPYVQNKEF IEVRLSSDID
 351 PIDSKTRVAL ATPILKKLS*

45 The cp7251 nucleotide sequence <SEQ ID 276> is:

1 ATGGCTCCAA TTCACGGAAG TAATGCGTTT GTTGAGGATA TTTTACATTC
 51 CCACCCCTCT CCACAAGCGA CTTATTTTTC TTCAACACGC GCCCAAAAAC
 101 TTCATGAGTT TAAAGACAGG CATCCCGTGC TTACACGGAT TGCTTCTGTA
 151 ATTATTAAAA TTTTAAAGT TCTGATAGGG CTGATCATCC TTCCCTTAGG
 201 AATCTACTGG CTATGFCAAA CGCTTTGTAC AAACCTCGATT CTCCCTTCCA
 251 AGAATTTATT AAAAATTTTC AAGAAGCAAC CCAACACTAA AACCTTAAAA
 301 ACTAATTATT TGCATGCTTT GCAAGATTAT TCCTCGAAAA ACCGCGTTGC
 351 TTCCATGAGA CGAGTTCCTA TCCTCCAGGA TAATGTCTTC ATCGACACTT
 401 TGGAAATATG CCTTTCACAA GCACCTACGA ATCGTTGGAT GCTCATTTCT
 55 451 TTAGGAAGTG ACTGTAGCTT GGAAGAAATC GCTTGTAAAG AGATCTTTGA

501 TTCTTGCCAA AGATTTGCCA AGTTGATAGG GGCCAATATA CTCGTTTATA
 551 ACTACCCCGG AGTCATGTCC AGCACAGGGA GCAGCAGCCT AAAGGACCTA
 601 GCATCAGCTC ATAATATTTG TACAAGATAC CTTAAAGATA AAGAACAGGG
 651 CCCTGGAGCA AAAGAAATCA TTACCTATGG GTACTCCCTA GGAGGTTTGA
 701 TACAAGCAGA AGCATTGCGA GACCAGAAGA TTGTTGCAAA CGATGATACT
 751 ACTTGGATAG CAGTCAAAGA TAGGTGTCCT CTCTTTATAT CTCCAGAAGG
 801 TTTCCACAGT TGCAGACGCA TAGGAAAGCT AGTAGCTCGT CTTTTTGGCT
 851 GGGGGACCAA AGCCGTAGAG AGAAGCCAAG ACCTTCCCTG CCTAGAAATT
 901 TTTCTCTATC CTACGGATTG CTTACGAAGA TCAACAGTCA GACAGAACAA
 951 GCTCTTAGCA CCTGAACCTA CTCTCGCTCA TGGGATAAAA AATAGTCCCT
 1001 ATGTTCAAAA TAAAGAATTT ATAGAAGTAC GATTATCGTC TGATATCGAT
 1051 CCCATCGACA GCAAAACAAG AGTGGCTCTT GCCACACCAA TTTTGAAAAA
 1101 GCTCTCTTAG

The PSORT algorithm predicts inner membrane (0.4545).

- 15 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 138A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 138B) and for FACS analysis.

These experiments show that cp7251 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

20 Example 139

The following *C.pneumoniae* protein (PID 4377288) was expressed <SEQ ID 277; cp7288>:

1 MHMSNPISLF SPAELIAKYN LIPKTSPIYP RRELIILEE NACQTRLTNV
 51 AQVLHPSSLF SMSKKILNPC GCSGGPLCWV ILNILAFIIT SVLFIILLPV
 101 NLIVAGLRLF MPLPPKKIVE DLSEPTTEET NEVIQPFIFA LQALLFEDNK
 151 LRSFKIVEQS VGKAPLPNPF LNRLVAISPQ ESQEA MRKIP DLCSQLKKVL
 201 KSLGVLTPFW KHM LKYFEG L KNEHDSNPDK KTFPILIKLL IEALTGKSSL
 251 PKTPSTKEKM QAALFIASSC KTCCKPTWGEV ITRSLNRLYS IANEGDNQLL
 301 IWVQEFKERE LMSIQDGD D AEEYRF AAQQH GERYTEAIEQ VLRNESAAKL
 351 QWHVINTMKF FHGKNLGLVT EHLQDTLGAL TLRQTTVDTH QGREDA DLSA
 401 ALFLNKYLS GNQLVNSVFK SMQKADPETK ALIREFALDI LYASLRLPQT
 451 SAHTEVFSTL LMDPETYEPN KACIAYLLYV LKIIEL*

The cp7288 nucleotide sequence <SEQ ID 278> is:

1 ATGCATATGT CTAACCCCAT CTCTTTGTTT TCCCTGCGAG AGTTAATAGC
 51 AAAGTACAAT TTAATTCCAA AAACCTCGCC GATTATATCCT CGGAGGACGG
 101 AACTTATTAT CTTGGAAGAA AATGCGTGTC AAACACGCCT AACCAACGTG
 151 GCTCAGGTCC TACATCCTTC TAGCCTATTC AGTATGTCAA AAAAAATACT
 201 GAATCCCTGC GGGTGCTCTG GTGGTCCCTT ATGTTGGGTG ATTCTCAACA
 251 TCCTAGCAT TATTATTACT TCAGTACTGT TTATCATTCT TTTACCGGTG
 301 AATCTCATCG TAGCAGGTCT TCGTCTCTTC ATGCCTCTTC CCCCTAAAAA
 351 AATCGTAGAG GATTTAAGTG AACCTACTAC TGAAGAAACG AATGAGGTCA
 401 TTCAACCCCTT CATTTTCGCT TTGCAAGCGT TGCTTTTTGA GGATAACAAA
 451 CTTGCTCTCT TTAATAATTGT TGAACAAAGT GTAGGCAAAG CACCCTTACC
 501 TAATCCCTTT TTAATAGAC TAGTAGCAAT TTCGCCGCAA GAAAGCCAAG
 551 AAGCCATGCG GAAGATTCCG GATCTATGCT CACAACCTGAA AAAAGTATTA
 601 AAGTCTCTAG GCGTGCTAAC TCCAGAATGG AAGCACATGC TGAAGTACTT
 651 TGAGGGACTG AAAAAACGAC ATGATAGTAA TCCTGATAAA AAGACGTTCC
 701 CAATATTGAT CAAGCTCCTC ATAGAAGCTC TTACTGGAAA GTCCCTCTTTA
 751 CCCAAAATC C'TAGTACAAA GGAAAAAATG CAAGCGGCCT TATTTATTGC
 801 AAGTTCTTGC AAGACTTGTA AGCCGACTTG GGGAGAAGTC ATAACCAGAT
 851 CTCTTAACAG ACTCTATAGT ATAGCTAATG AAGGAGACAA TCAGCTTCTG
 901 ATTGGGGTTC AAGAGTTTAA AGAACGAGAG CTGATGTCCA TCCAAGATGG
 951 TGATGATGCT GAAGAGTATC GGTTTGCGGC TCAGCAACAC GGTGAGCGTT
 1001 ACACAGAGGC AATAGAACAA GTTCTACGAA ACGAGTCAGC AGCCAAACTA
 1051 CAATGGCATG TGATCAACAC TATGAAATTC TTCCATGGGA AAAATCTCGG
 1101 TCTAGTTACA GAACACCTAC AAGATACTCT CGGCGCCTA ACTTTACGTC
 1151 AACTACAGT GGACACACAT CAAGGCAGAG AAGACGCTGA TTTGTCAGCT
 1201 GCTCTTTTCC TAAATAAGTA TTTAAATTTCT GGAAATCAAC TTGTTAATAG

1251 CGTCTTTAAA TCCATGCAAA AAGCAGATCC AGAAACCAAA GCTTTAATCC
 1301 GTGAGTTTGC TCTAGATATA TTATATGCAT CCTTACGGCT TCCTCAAACT
 1351 TCCGCTCATA CCGAGGTCTT TTCTACACTC TTAATGGACC CAGAGACCTA
 1401 TGAACCTAAT AAAGCTTGTA TCGCCTACTT GCTCTATGTA TTAAAGATCA
 1451 TCGAATATA A

The PSORT algorithm predicts inner membrane (0.5989).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 139A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 139B) and for FACS analysis.

10 These experiments show that cp7288 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 140

The following *C.pneumoniae* protein (PID 4377359) was expressed <SEQ ID 279; cp7359>:

15 1 MPGSVSSPPL SPVIVRERV SSSGSDLIQP HAVLKISILI FALVTILGIV
 51 LVLVSSALGA LPSLVLTVSG CIAIAVGLIG LGILVTRLIL STIRKVDAMG
 101 YDAVKKEQY LSRIRELESE NREIRDRNRA VEDQCAHLSE ENKDLRDPEY
 151 LHGMTERLIA SLEIENQALV AENILLKQDWN ASLSRDFRAY KQKFPLGALE
 201 PWKEDIACIM EQNLFLKPEC IAMVKSLEPLE TQRLFLYPKG FQSLVNRFPAP
 251 RSRPFQTPKY EYNSRNENED GKVAAVCARL KKEFFSAVLG ACSYEELGGI
 301 CERAVALKET LPLPEAVYDT LVQEFFNLIT AESLWKEWCF YSYPYLRPYL
 351 SVDYCKRLFV QLFELCLKL FTTGSPEDQA LVRLFSSYRN HIPAVLASFG
 401 LPPPETGGSV FVLLPKQENL LWSQIEVLAT RYLKDTFVRN SEWTGSFEMM
 451 FSYNEMCKEI SEGRIRFABD YETRHSEEPF PSPLSEEGEG EEFLLPCCSE
 501 EVSVLERPD LVDSDMWVHP PVPKGPL*

25 The cp7359 nucleotide sequence <SEQ ID 280> is:

1 ATGCCAGGTT CTGTGTCATC ACCTCCTTTG TCTCCTGTAA TTGTCCGTGA
 51 AAGGGTCCCA TCCTCTTCAG GATCCGACCT CATACAGCCT CATGCTGTCTT
 101 TAAAGATCTC CATCCTAATT TTTCGCGCTG TGACAATTTT AGGAATTGTT
 151 CTGTAGTGT TGTCTAGTGC TTTAGGAGCT CTTCCTAGTT TAGTTTTCGAC
 201 GGTTCCTGGT TGTATTGCAA TAGCTGTAGG CCTGATTGGT TTAGGGATTG
 251 TTGTGACACG GCTGATTCTC TCTACGATCA GAAAAGTAGA TGCCATGGGT
 301 TATGATGCTG CGGTCAAAGA AGAGCAGTAT TTGTCACTGA TCAGAGAATT
 351 AGAGTCTGAA AATAGAGAGA TTAGAGATAG AAATCGTGCT GTCGAAGATC
 401 AGTGTGCCCA TTTATCCGAA GAGAACAAGG ACCTTAGGGA TCCCGAATTAT
 451 CTACATGGAA TGACTGAAAG GCTCATTGCG AGCTTAGAAA TAGAGAATCA
 501 AGCTCTCGTA GCTGAGAACA TTCTTCTCAA AGACTGGAAT GCAAGCCTAT
 551 CTAGAGATTI CCGCGCATAT AAGCAAAAAT TTCCTCTTGG GGCATTAGAA
 601 CCCTGGAAAG AAGATATTGC ATGTATCATG GAACAAAATC TCTTTTAAAA
 651 ACCGGAATGT ATCGCGATGG TTAAGTCTCT TCCATTAGAG ACGCAACGGC
 701 TGTTTTTATA TCCAAAAGGA TTTCAGTCTT TAGTTAATCG ATTTGCTCCG
 751 CGGTCTCGCT TTTTCCAGAC TCCAAAGTAT GAATATAACA GTAGGAATGA
 801 AAATGAGGAC GGAAAGGTAG CCGCAGTGTG CGCCCGTTTG AAAAAAGAAT
 851 TCTTCAGTGC TGTTTTCAGG GCTGTAGTT ACGAAGAACT AGGGGGCATT
 901 TGTGAAAGAG CAGTAGCACT TAAAGAGACG TTGCCATTGC CTGAAGCTGT
 951 CTATGATACC CTAGTTCAGG AGTTCCCAAA TCTTCTTACT GCTGAGAGTT
 1001 TATGAAAGA ATGGTGCTTC TATTCTATC CCTACCTTCG TCCCTATCTT
 1051 TCTGTGGATT ACTGTAAGAG GTTATTGTGA CAACTTTTGG AGGAACCTCG
 1101 CCTAAAGCTT TTTACAACGG GATCTCCAGA AGACCAAGCT TTGGTTTCGCC
 1151 TTTTCTCTTA CTATAGGAAT CATATTCCCG CAGTCTTGGC CTCATTGGGT
 1201 TTGCCCCCGC CTGAGACAGG GGGGTCTGTA TTTGTATTGC TACCAAAACA
 1251 AGAAAACCTT CTTTGAGATC AAATTGAGGT GCTGGCTACA AGGTATCTCA
 1301 AAGATACCTT CGTGAGAAAC TCAGAAATGA CGGGCTCTTT CGAGATGATG
 1351 TTTTCTTATA ACGAGATGTG TAAGGAGATC TCCGAAGGAA GGATTTCGTTT
 1401 TGCTGAAGAC TATGAAACGA GGCATTCCGA AGAATTCCTT CCTTCCCTC
 1451 TCTCTGAAGA AGGAGAGGGC GAAGAATTCC TTCCTCCTTG CTCTGAAGAA
 1501 GAGGTTTCGG TTCTTGAGCG CCCAGATCTA GATGTAGACT CTATGTGGGT
 1551 CTGGCATCCG CCGGTCCCTA AGGACCTCTT TTAA

The PSORT algorithm predicts inner membrane (0.7453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 140A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 140B) and for FACS analysis.

- 5 These experiments show that cp7359 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 141

The following *C.pneumoniae* protein (PID 4377374) was expressed <SEQ ID 281; cp7374>:

```

10      1 MDKQSSGNSG CIWHPFTQSA LDSTPIKIVR GEGAYLYAES GTRYLDIAISS
      51 WWCNHLHGHH PYITKKLCEQ AQKLEHVIFA NPTHEPALEL VSKLAPLLPE
     101 GLERFFFSND GSTSIEIAMK IAVQYYYNQN KAKSHFVGLS NAYHGDTFGA
     151 MSIAGTSPTT VPFHDLFLPS STIAAPYYGK EELAIQAQKT VFSIESNIAAF
     201 IYEPLLOQAG GMLMYNPEGL KEILKLAKHY GVLCIADEIL TGFGRGTPLF
     251 ASEFTDIPPD IICLSKGLTG GYLPLALTVT TKEIHDAFVS QDRMKALLHG
     301 HTFTGNPLGC SAALASLDLT LSPECLQQRQ MIERCHQEFQ EAHGSLWQRC
     351 EVLGTVLALD YPAEATGYFS QYRDHLNRFF LERGVLLRPL GNTLYVLPFY
     401 CIQEEDLRIT YSHLQDALCL QPQ*
```

The cp7374 nucleotide sequence <SEQ ID 282> is:

```

20      1 ATGGACAAGC AATCATCAGG GAATTCAGGG TGTATCTGGC ACCCCTTCAC
      51 TCAATCTGCA TTAGATTCTA CACCCATAAA GATTGTAAGG GGAGAAGGTG
     101 CTTACCTCTA TGC CGAATCA GGAACAAGAT ATCTTGATGC GATATCTPCA
     151 TGGTGGTGCA ACCTCCACGG TCATGGGCAT CCCTACATTA CAAAAAATT
     201 ATGTGAGCAA GCACAGAAGT TAGAACATGT GATCTTCGCA AATTTACCCC
     251 ATGAACCGGC TCTAGAGCTC GTATCGAAAC TCGCTCCCTT CCTTCCTGAA
     301 GGTCTAGAAC GTTCTCTTTT CTCTGACAAC GGATCAACGT CTATCGAAAT
     351 AGCAATGAAA ATTGCTGTGC AATATTACTA CAATCAAAAC AAGGCTAAGA
     401 GCCATTTTGT TGGACTCAGC AATGCCTATC ACGGAGATAC ATTTGGAGCT
     451 ATGTCGATAG CTGGCAGGAG CCCTACTACA GTTCCCTTTC ATGATCTTTT
     501 TCTTCCCTTC AGTACAAATTG CTGCTCCCTA TTATGGCAAG GAAGAGCTTG
     551 CCATTGCCCA AGCAAAAACA GTCTTTTCTG AAAGCAATAT CGCAGCGTTT
     601 ATCTATGAGC CGCTATTGCA AGGTGCTGGA GGGATGTTAA TGTATAATCC
     651 CGAAGGCCTA AAGGAGATTC TCAAGCTTGC CAAGCATTAC GGGGTCTCTT
     701 GTATTGCTGA TGAATTTCTT ACTGGCTTTG GCCGTACGGG TCCACTGTTT
     751 GCTTCTGAAT TTACAGACAT TCCTCCTGAC ATTATCTGTC TTTCTAAAGG
     801 TCTTACAGGA GGCTATCTCC CTCTAGCCTT GACAGTAACC ACTAAAGAAA
     851 TTCATGATGC CTTTGTCTCC CAAGATCGGA TGAAGGCACT GCTTCATGGC
     901 CATACTTCA CAGGAAATCC TTTAGGCTGT AGTGCTGCCC TCGCTTCTTT
     951 GGATCTCACC CTATCTCCAG AATGCCTACA ACAAGGCAA ATGATAGAAC
    1001 GGTGTCATCA AGAGTTTCAA GAAGCTCATG GTTCCCTATG GCAACGGTGT
    1051 GAGGTTCTGG GCACGGTACT CGCTCTAGAT TACCCTGCAG AAGCTACAGG
    1101 ATATTTTTC CAATATAGAG ACCATCTCAA TCGCTTTTTC TTAGAACGTG
    1151 GAGTCCTTCT TCGTCTTTTA GGAACACAC TGTATGTGCT GCCCCCTTAC
    1201 TGTATCCAAG AAGAAGATCT CCGGATTATT TATTCTCACC TACAGGATGC
    1251 CCTATGTCTA CAACCACAGT AA
```

- 45 The PSORT algorithm predicts cytoplasm (0.2930).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 141A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 141B) and for FACS analysis.

- 50 These experiments show that cp7374 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 142

The following *C.pneumoniae* protein (PID 4377377) was expressed <SEQ ID 283; cp7377>:

```

1 MREETVSWSL EDIREIYHTP VFELIHKANA ILRSNFLHSE LQTCYLISIK
51 TGGCVEDCAY CAQSSRYHTH VTPEPMMKIV DVVERAKRAV ELGATRVCLG
101 AAWRNAKDDR YFDRVLAMVK SITDLGAEVC CALGMLSEEQ AKKLYDAGLY
151 AYNHNLDSSE EFYETIITR SYEDRLNLTLD VVNKSGISTC CGGIVMGES
201 EEDRIKLLHV LATRDHIPS VFNLLWPID GTPLQDQPP I SFWEVLRTIA
251 TARVVFPRSM VRLAAGRAFL TVEQQTL CFL AGANSIFYGD KLLTVENNDI
301 DEDAEMIKLL GLIPRPSFI EGNPCYANN S*

```

10 The cp7377 nucleotide sequence <SEQ ID 284> is:

```

1 ATGCGTGAAG AAAGTGTATC CTGGTCATTA GAAGACATCC GCGAAATTTA
51 TCACACTCCC GTATTGAGC TGATTCACAA AGCCAATGCC ATATTGCGTA
101 GTAATTTCCT CCATTCAGAA CTGCAGACTT GCTATCTGAT TTCGATTAAA
151 ACTGGTGGAT GCGTTGAAGA TTGCGCTTAC TGTGCCCAAT CTTCGCCGTA
201 TCATACCCAC GTCACACCAG AACCTATGAT GAAAATTGTA GACGTTGTGG
251 AAAGGCAAAA ACGTGTCTGTA GAGCTAGGCG CCACTCGTGT GTGTCTTGGG
301 GCTGCCTGGC GCAATGCTAA GGACGATCGA TACTTTGATA GAGTCTTCGC
351 TATGGTGAAA AGTATCACAG ATCTCGGAGC CGAGGTTTGT TGTGCTTTAG
401 GCATGCTCTC CGAAGAGCAA GCTAAAAAAC TGTATGATGC AGGACTTTAT
20 451 GCCTACAATC ATAATTAGA CTCTTCTCCG GAATTCTATG AAACATAAAT
501 CACAACACGT TCTTATGAAG ATCGCCTCAA CACTCTTGAT GTAGTAAATA
551 AATCTGGCAT TAGTACATGC TGCCTGGTGA TTGTAGGTAT GGGAGAATCT
601 GAAGAAGACC GTATAAAGCT TCTTCATGTT CTTGCAACAA GAGATCATAT
25 651 CCCAGAATCC GTACCTGTAA ATTTACTTTG GCCGATTGAC GGCACGCCTT
701 TGCAAGACCA GCCTCCGATT TCTTCTGGG AAGTCTTGGC AACCATAGCA
751 ACGGCACGGG TTGTTTTCCC CAGATCCATG GTACGACTTG CTGCAGGACG
801 CGCTTTCCCT ACAGTAGAAC AACAAACCTT ATGTTTTCTA GCCCGTGCCA
851 ACTCCATATT CTATGGAGAT AAACGTGTTGA CTGTAGAAAA CAATGATATA
901 GATGAAGATG CTGAAATGAT CAACTTTTA GGCTTAATCC CTCGCCCTTC
30 951 ATTTGGAATA GAAAGAGGTA ACCCATGTTA TGCCAACAAT TCCTAA

```

The PSORT algorithm predicts cytoplasm (0.2926).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 142A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 142B) and for FACS analysis.

35 These experiments show that cp7377 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 143

The following *C.pneumoniae* protein (PID 4377407) was expressed <SEQ ID 285; cp7407>:

```

40 1 MVEPNNSWFR MCGNFNCEWV EVTTTEETTR QSASDISEEA GSSGGAAPIT
51 TQPTKITKVE KRVQFNATAQ DESTIHIQE AGELVDSILS HRRTQGCTEY
101 CYDSYATGCG QRCGSFGRLL CGTYKACCLD REDNQVAGLV HECEQTHGPI
151 AVALAAKTMG LNLMEVLEKN TILSEEQKNE FRQHCSEAKT QLYGTMQSLS
201 QNFFLEGVNS IREGLDDSL VQAVLSFIAT RSWERTIESE EASGTSSASN
45 251 STRIPACYIL NTSPLTTSRL SCGSRDARRP SSVGAEPQYV AKKYNDNGMA
301 RQLGKIQVTN LKTGDFSAIG PFGLLIVKML NSFLLSASQS TSSILKHTGG
351 EICYTCPNFR DIVLLMLAI GYCPANTDET SVVDIHMIDD PIMTIFYRLQ
401 YSYRTGKTS SFLKKKPSLV RQESLDCPTP AESVPLMSSL EEDENEDDD
451 EDGNLAYQQR ILECSGHLQT LFLGIKINKE *

```

The cp7407 nucleotide sequence <SEQ ID 286> is:

```

50 1 ATGGTTTGCC CAAATAATTC TTGGTTCAGA ATGTCTGGAA ATTTCAACTG
51 CGAATGGGTT GAAGTAACAA CAACAGAAGA AACACGCGG CAATCGGCTT
101 CAGATATAAG CGAAGAAGCT GTTCGAGTG GAGGAGCTGC TCCTATAACT
151 ACGCAACCTA CTAATAATTAC AAAAGTAGAG AAACGTGTCC AATTTAATAC

```

201 TGCTCAAGGT GATGAAAGTA CAATACACAT GATCCAAGAA GCAGGAGAAT
 251 TGGTAGACTC CATTCTATCA CATAGACGAA CGCAAGGATG TACAGAGTAT
 301 TGTATGACA GTTACGCAAC TGGATGTGGT CAGCGTTGCG GATCTTTTGG
 351 AAGACTCAT TGTGGAACGT ATAAAGCGTG TTGCTTAGAC AGAGAGGATA
 401 ATCAGGTTGC TGGACTTGTC CATGAATGCG AACAGACCCA TGGTCCTATT
 451 GCCGTTGCTT TAGCTGCTAA AACTATGGGC CTCAACTTAA TGGAACTTGT
 501 AGAAAAAAC ACTATTTTGT CTGAAGAACA GAAAAATGAA TTTAGACAGC
 551 ATTGCTCGGA AGCTAAACC CAACTCTATG GAACGATGCA GAGCCTTTCT
 601 CAAAACCTTT TCCTTGAAGG AGTCAACAGC ATTAGAGAAC GCGTCTAGA
 651 CGATTCAC TA GTCCAAGCCG TGCTAAGCTT TATTGCTACA AGGTCTTGGG
 701 AAAAACTAT AGAATCAGAG GAAGCCTCAG GAACATCTTC TGCTTCTAAT
 751 TCTACACGCA TPCCTGCGTG CTATATCTTA AATACGAGCC CCTTAACGAC
 801 GTCACGCTA TCCTGTGGAT CAAGAGATGC GCGACGCCCA TCTTCAGTCG
 851 GTGCAGAGCC CCAGTACGTA GCAAAAAAAT ACAATGACAA TGGCATGGCC
 901 AGACAAATAG GAAAAATCCA AGTCACCAAT CTAAAAACAG GAGATTTTTC
 951 AGCTTTAGGT CCTTTTGGTC TCCTGATGTG GAAAAATGCTG AATAGCTTTC
 1001 TCTTATCTGC ATCACAAGC ACATCTTCTA TTCTAAAGCA CACAGGTGGA
 1051 GAAATATGTT ATACGTGCCC AAATTTTCGT GATATCGTCG TTTTATTGAT
 1101 GTTAGCGATT GGCTATTGCC CTGCAAATAC CGATGAGACA TCTGTCGTAG
 1151 ATATACACAT GATAGATGAT CCGATTATGA CCATCTTCTA TCGACTACAA
 1201 TACAGCTATA GAACAGGGA AACTTCAGCA TCGTTTTTAA AAAAGAAACC
 1251 CTCATTAGTA AGACAGGAAA GTCTTGATTG TCCTACCCCT GCAGAACTCG
 1301 TCCCTCTCAT GTCTGCTC GAAGAAGAAG ATGAAAATGA AGATGATGAT
 1351 GAGGATGGGA ATTTGGCGTA TCAACAGCGT ATCCTTGAAT GCTCGGGTCA
 1401 TTTACAACT CTATTTTATG GGATAAAAT AAACAAGAA TAA

The PSORT algorithm predicts inner membrane (0.1319).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 143A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 143B) and for FACS analysis.

- 30 These experiments show that cp7407 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone:

Example 144

The following *C.pneumoniae* protein (PID 4376432) was expressed <SEQ ID 287; cp6432>:

1 MTRSTIESSD SLCSRSFSQK LSVQTLKNLC ESRLMKITSL VIAFLTLIVG
 51 GALIALAGGG VLSFPLGLIL GSVLVLFSSI YLVSCCKFFT LKEMTMTCV
 101 KSKINWFEK QRNKDIEKAL ENPDLFGENK RNVGNRSARN QLEMILHETD
 151 GIILKRYMKG AKMYFYL*

The cp6432 nucleotide sequence <SEQ ID 288> is:

1 ATGACTAGAA GTACTATTGA AAGCAGTGAT TCGCTATGCT CAAGGTCTTT
 51 TTCTCAAAAA TTAAGTGTCC AGACATTAAA AAATCTCTGT GAAAGTAGAT
 101 TAATGAAGAT CACTTCTCTT GTGATTGCTT TCCTAACTCT AATTGTGGGG
 151 GGTGCTCTTA TAGCTTTAGC AGGAGGGGGG GTTCTTTCTT TCCCTCTTGG
 201 GCTAATCTTA GGAAGCGTAC TCGTTTGTG TTCTTCTATC TATTTAGTCT
 251 CTGTTGTAA ATTTTCTACT TTAAAAGAGA TGACAATGAC CTGTAGTGTC
 301 AAATCTAAAA TCAATATATG GTTTGAAAAG CAACGAAACA AAGACATCGA
 351 AAAGGCATTA GAGAATCCAG ATCTCTTTGG AGAAAATAAG AGAAATGTTG
 401 GAAATCGTTC GGCAAGAAAT CAACTAGAAA TGATCTTACA CGAGACTGAC
 451 GGAATTATTT TGAAAAGATA TATGAAAGGA GCTAAAATGT ACTTTTATTT
 501 ATGA

- 50 The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 144A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 144B) and for FACS analysis.

These experiments show that cp6432 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 145

The following *C.pneumoniae* protein (PID 4376433) was expressed <SEQ ID 289; cp6433>:

```

5      1  MNWVPKTI DH VDPESIDIR KVVSCYKLIK ECQPEFRSLI SELLGVIRCG
      51  LRLLRKRSKYQ EQARTVSDSD APLFCLTRSY YQDGYLTPLR AGPRDLINHY
     101  IHLRRRENPK HFFSPKHPY YARLAFNESV CVYRELFIDIE RLTKMYVEGD
     151  YSKEQEKNLQ AILSFVKTL D EGKDFLIEHK DTDLIGRGFT DVFCT*
  
```

The cp6433 nucleotide sequence <SEQ ID 290> is:

```

10      1  ATGAATTGGG TTCCAAAAAC AATAGACCAT GTAGATCCAG AATCAGAGAT
      51  AGATATACGT AAAGTCGTCT CCTGCTATAA GTTGATAAAA GAATGTCAAC
     101  CTGAATTTCG ATCTCTTATA AGTGAATTAC TAGGAGTGAT TCGGTGTGGC
     151  TTAAGACTAT TAAAACGTTT TAAGTATCAA GAACAGGCTA GAACTGTATC
     201  TGATGAAGAT GCACCTCTTT TCTGCCTGAC TCGTCTTTAT TATCAAGATG
     251  GTTATCTCAC GCCATTAAAG GCAGGACCTC GTGATCTTAT AAATCACTAT
     301  ATACACTTGC GTCGCCGAGA GAATCCTAAG CATTTTTCFA GTCTTAAGCA
     351  TCCATGTTAT TATGCTCGAT TGGCTTTTAA TGAGTCAGTG TGTGCTCTATA
     401  GAGAACTCTT TGATATAGAG CGACTTACAA AAATGTATGT CGAGGGTGAT
     451  TATTCTAAAG AACAAAGAGAA AACCTACAG GCTATTCTTA GTTTGTGTGA
     501  AACTCTAGAT GAAGGAAAGG ACTTCTTTAT TGAACATAAA GATACCGATC
     551  TCATTGGGAG AGGTTTTACT GATGTGTTCT GCACTTAA
  
```

The PSORT algorithm predicts cytoplasm (0.4068).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 145A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 145B) and for FACS analysis.

These experiments show that cp6433 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 146

The following *C.pneumoniae* protein (PID 4376643) was expressed <SEQ ID 291; cp6643>:

```

30      1  MGYLPVSATD VLFESPAAPL INSANTQNQK LIELKGKQQA ESSPRTITSV
      51  ILEVLLVIGC CLIVLSLLAI RPALQFTLET GHFAAIAVLA VSGTILLVAV
     101  IILFCFLAAV PFAAKKTYKY VKTVDDYASW HSHQQTPTLG TIFSGIVYAE
     151  SQAQL*
  
```

The cp6643 nucleotide sequence <SEQ ID 292> is:

```

35      1  ATGGGATATC TTCCAGTATC TGCTACGGAC GTTCTTTTGT AAAGTCCAGC
      51  CGCTCCCTTA ATCAATAGCG CAAACACACA AAATCAGAAA CTCATAGAAC
     101  TCAAGGGGAA GCAGCAAGCT GAGTCTTCTC CACGGACAAT CACTTCTGTC
     151  ATATTGGAAG TTCTCCTAGT GATCGGATGC TGCCTCATAG TTCTTAGTTT
     201  ATTGGCAATC CGCCCTGCTC TGCAATTAC TCTAGAACT GGACATCCAG
     251  CTGCCATTGC AGTCCCTTGT GTCTCAGGAA CAATTCTATT GGTGGCTGTT
     301  ATCATCTTGT TTGCTTTCT AGCAGCTGTG CCATTGCTG CTAAGAAAAC
     351  TTATAAATAT GTTAAGACGG TTGATGACTA TGCTTCTTGG CATTCTCATC
     401  AGCAAACACC GACCTAGGC ACTATCTTTT CAGGTATCGT CTATGCAGAA
     451  TCCAGGCGC AATTATAG
  
```

The PSORT algorithm predicts inner membrane (0.6859).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 146A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 146B) and for FACS analysis.

These experiments show that cp6643 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 147

The following *C.pneumoniae* protein (PID 4376722) was expressed <SEQ ID 293; cp6722>:

```

1  VSSTLNGVFP SSLPEESADL FITNKEIVAL GEKGNVFLTH SIPMHIAAIT
51  ILVIVALAGI AIICLCGYSQ SILLIAVGIV LTILTLLCLQ ALVGFIFKIR
101 QLPQQLHTTV QFIREKIRPE SSLQLVTNAQ RKTQTDLKL YEELCDLSQK
151 EFKLQSTLYQ KRFELSHKNE KTNQN*

```

The cp6722 nucleotide sequence <SEQ ID 294> is:

```

1  GTGTCTAGTA CTTTAAACGG GGTATTTCCTC TCATCCCTTC CGGAAGAGTC
51  TGCTGATTTA TTCAATTACGA ATAAGGAGAT CGTAGCTTTG GGGGAGAAGG
151 GCAATGTTTT TCTCACCAC TCCATTCCTA TGCATATTGC TCGGATTACG
151 ATCTTAGTGA TTGTAGCTCT TGCTGGAATC GCTATTATCT GTTTGGGTTG
201 CTATAGCCAA AGCATTCGTG TGATTGCCGT TGGCATTGTT CTTACTATTT
251 TGACTCTTCT CTGCCACAA GCCTTGGTAG GATTATTAA ATTCATCCGG
301 CAGCTCCCTC AGCAGCTCCA TACGACAGTA CAATTTATCA GGGAGAAGAT
201 TCGACCTGAA TCCTCTCTAC AGCTTGTAAC CAATGCACAG AGAAAAACCA
401 CTCAAGATAC GCTAAAGTTA TACGAAGAAC TCTGCGACCT CTCACAAAAA
451 GAGTTCAAAC TGCAATCAAC TCTTTATCAA AAACGTTTGT AGCTTTCTCA
501 CAAGAATGAA AAGACAAATC AAAACTAG

```

The PSORT algorithm predicts inner membrane (0.6668).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 147A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 147B) and for FACS analysis.

These experiments show that cp6722 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 148

The following *C.pneumoniae* protein (PID 4377253) was expressed <SEQ ID 295; cp7253>:

```

1  MSELAPCSTG LQMPVHTQVH HALDTRRVIL TIAACLSLIA GIVLVGLGAA
51  AILPSLFGVI GGMILILFSS IALYLYKKT REVDQIALEP LPEMISKDQS
101 IIDFVKTRDY ASLEKKATFA YTHTHYYDGS MVFYREIPRF MLGSYLALRK
151 DMDRQALF*

```

The cp7253 nucleotide sequence <SEQ ID 296> is:

```

1  ATGAGCGAGC TCGCCCCCTG CTCGACAGGA TTGCAGATGG TCCCCCATACT
51  GCAGGTCCAT CATGCCCTTG ATACGCGGAG AGTCATTCTA ACGATAGCCG
101 CCTGTCTGTC TTTAATTGCA GGAATCGTGT TGGTTGGCTT AGGTGCTGCA
151 GCAATCCTGC CCTCGCTTTT TGGAGTCATT GGAGGAATGA TTCTTATTCT
201 GTTTTCTTCG ATCGCCCTCA TTTATTTATA CAAGAAGACA AGGGAGGTGG
251 ATCAGATTGC TCTGGAGCCT CTTCTGAGA TGATTCTTAA AGATCAAAGC
301 ATTATAGATT TTGTAAAGAC ACGAGACTAT GCATCTTTAG AAAAGAAAGC
351 GACCTTTGCT TATACTCATA CTCATTATTA CGATGGAAGC ATGGTCTTCT
401 ATAGGGAGAT CCCTAGATTT ATGTTAGGCT CTTATCTCGC GCTTCGCAAA
451 GACATGGACC GCCAAGCTCT TTTTGA

```

The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 148A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 148B) and for FACS analysis.

These experiments show that cp7253 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 149

The following *C.pneumoniae* protein (PID 4376264) was expressed <SEQ ID 297; cp6264>:

```

1  VISGLLFLLV RREVPTVRSE EIPRGVSVTP SEEPALEKAQ KEPETKKILD
51  RLPKELDQLD TYIQEVFACL ERLKDPKYED RGLLTEAKEK LRVFDVVEKD
101 MMSEFLDIQR VLNEEAYYVE HCQDPLENIA YEIFSSQELR DYYCAGVCGY
151 LPSGDARADR LKRSVKEVMD RFMRVTWKS EASVMLDHSY GVARELFKKA
201 VGVLEESVYK ILFKSYRDAF YECEKAKIQR DGRFKWL*

```

The cp6264 nucleotide sequence <SEQ ID 298> is:

```

1  GTGATTTCGG GACTTCTATT CCTTCTAGTA AGACGAGAGG TTCCGACAGT
51  ACGTTCAGAG GAAATTCCCA GAGGGGTTTC TGTGACCCCT TCTGAAGAGC
101 CTGCTCTAGA GAAGGCTCAA AAAGAACCGG AGACAAAGAA AATTTTAGAT
151 CGGTTGCCGA AGGAATTGGA TCAGTTAGAT ACGTATATTC AGGAAGTGTT
201 TGCATGTTTA GAGAGGCTGA AGGATCCTAA GTACGAAGAT CGAGGTCTTT
251 TAACAGAGGC GAAGGAGAAA CTTGAGTTT TTGACGTTGT TGAGAAAGAT
301 ATGATGTCAG AGTTTTTAGA CATAACAAGA GTGTTGAATG AGGAAGCATA
351 TTATGTAGAA CATTGTCAAG ATCCCCTAGA GAATATAGCC TACGAGATTT
401 TCTCTTCCCA AGAGCTTCGT GATTACTACT GTGCAGGGGT GTGTGGGTAT
451 TTGCCTTCTG GGGATGCTCG AGCGGATCGA TTAAGAGAT CAGTTAAGGA
501 GGTAAATGGAT CGCTTTATGA GGGTGACCTG GAAATCTTGG GAGGCATCAG
551 TCATGTTGGA TCATAGCTAT GGGGTAGCGC GAGAGTTATT CAAGAAGGCA
601 GTAGGAGTAC TAGAGGAGAG TGTCTATAAA ATTCTGTTTA AGAGCTATAG
651 AGATGCGTTT TATGAATGTG AGAAGGCAAA GATCCAGAGG GATGGCGGTT
701 TCAAATGCTT ATAG

```

The PSORT algorithm predicts cytoplasm (0.2817).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 149A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 149B) and for FACS analysis.

These experiments show that cp6264 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 150

The following *C.pneumoniae* protein (PID 4376266) was expressed <SEQ ID 299; cp6266>:

```

1  MLLISGALF LTLGIPGLSA AISFGLGIGL SALGGVLMIS GLLCLLVKRE
51  IPTVRPEEIP EGVSLAPSEE PALQAAQKTL AQLPKELDQL DTDIQEVFAC
101 LRKLKDSKYE SRSFLNDAKK ELRVDFVVE DTLSEIFBLR QIVAQEGWDL
151 NPLINGGRSL MMTAESESLD LFHVSRLGY LPSGDVRGEG LKKSKEIVA
201 RLMSLHCEIH KVAFAFRNS YAMAEKAFK ALGALEESVY RSLTQSYRDK
251 FLESERAKIP WNGHITWLRD DAKSGCAEKK LGMPRNVGRN LGKQSFSG*

```

The cp6266 nucleotide sequence <SEQ ID 300> is:

```

1  ATGCTCTTAC TGATTTCAGG AGCTCTCTTT CTGACGTTAG GGATTCCAGG
51  ATTGAGTGCA GCAATTTCTT TTGGATTAGG CATCGGTCTC TCCGCATTAG
101 GAGGAGTGCT GATGATTTCG GGACTACTAT GTC'TTTTAGT AAAACGAGAG
151 ATTCCGACAG TACGACCAGA AGAAATTCCT GAAGGGGTTT CGCTGGCTCC

```

201 TTCTGAGGAG CCAGCTCTAC AGGCAGCTCA GAAGACTTTA GCTCAGCTGC
 251 CTAAGGAATT GGATCAGTTA GATACAGATA TTCAGGAAAGT GTTCGCATGT
 301 TTAAGAAAGC TGAAAGATTG TAAGTATGAA AGTCGAAGTT TTTTAAACGA
 351 TGCTAAGAAG GAGCTTCGAG TTTTGTGACTT TGTGGTTGAG GATACCTCT
 401 CGGAGATTTT CGAGTTGCGG CAGATTGTGG CTCAAGAGGG ATGGGATTTA
 451 AACTTTTGA TCAATGGGGG ACGAAGCCTC ATGATGACTG CAGAATCTGA
 501 ATCGCTTGAT TTGTTCATG TATCGAAGCG GCTAGGGTAT TTACCTTCTG
 551 GGGATGTTTC AGGGGAGGGG TTAAGAAAT CTGCGAAGGA GATAGTCGCT
 601 CGTTTGATGA GCTTGCATTG CGAGATTCAC AAGGTGGCGG TAGCGTTTGA
 651 TAGGAATTCC TATGCGATGG CAGAAAAGGC GTTTGCGAAA GCGTTGGGAG
 701 CTTTAGAAGA GAGTGTGTAT CGGAGTCTGA CGCAGAGTTA TAGAGATAAA
 751 TTTTGGAGA GCGAGAGGGC GAAGATCCCA TGGAAATGGC ATATAACCTG
 801 GTTAAGAGAT GATGCGAAGA GTGGGTGTGC TGAAGAAGAG CTGGGGATGC
 851 CGAGGAACGT TGAAGAAAT TTAGGAAAGC AGTCTTTTGG GTAG

15 The PSORT algorithm predicts inner membrane (0.3590).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 150A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 150) and for FACS analysis.

20 These experiments show that cp6266 is a surface-exposed and immunoaccessible protein and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 151

The following *C.pneumoniae* protein (PID 4376895) was expressed <SEQ ID 301; cp6895>:

1 MKIKKSFQYS LCQAKRFQNM LPNHFDPCIQ PVNLQLKQDR LAYGELIILL
 25 51 SKYQQRTFSS LLKEETCSLN RAKQHLLYKI LRDFNTMQHL RSLGLNGWGE
 101 IFMSPCL*

The cp6895 nucleotide sequence <SEQ ID 302> is:

1 ATGAAGATTA AAAAATCTTT TCAATACAGT TTATGCCAAG CAAAGAGATT
 51 TCAGAACATG CTGCCAAACC ACTTTGATCC ATGTTTGCAG CCAGTGAATT
 101 TACAACCTCAA ACAAGACAGA TTGGCATAAG GGGAGCTCAT CATATTGCTA
 151 TCTAAATATC AACAAAAGAC CTTTTCCTCT TTGTTGAAGG AAGAAACATG
 201 TTCTCTTAAT CGTGCGAAGC AGCACTTATT GTATAAGATT TTGAGAGATT
 251 TTAATACTAT GCAGCATCTA AGGTCCCTCG GATTAAATGG TTGGGGAGAG
 301 ATCCCTATGA GTCCTTGCTT CTAA

The PSORT algorithm predicts cytoplasm (0.3264).

35 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 151A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 151B) and for FACS analysis.

These experiments show that cp6895 is a surface-exposed and immunoaccessible protein and that it is a useful immunogen. These properties are not evident from the sequence alone.

40 Example 152 and Example 153

The following *C.pneumoniae* protein (PID 4376282) was expressed <SEQ ID 303; cp6282>:

1 MSLNLPPSSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
 51 KLNYPKKLII IEKELKTLFP LLMRKGTLP KRRPDILIT PPTYTDAQGN
 101 THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
 151 ALFNPKTQTL DFYPGLPEYS QLLNYFISLN L*

The cp6282 nucleotide sequence <SEQ ID 304> is:

```

1  ATGTCCTTAT TGAACCTTCC CTCAAGCCAG GATTCCTGCAT CTGAGGACTC
51 CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
101 CTACTCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
151 AAGCTGAACCT ACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAAC
201 TCTTTTCCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
251 CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
301 ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
351 CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
401 CTATCGGAGC CACCTGCATT GCTATGGCAG GGAAACACTC TCAAGTGTCA
451 GCTCTCTTCA ATCCAAAAAC ACAACTCTT GATTTTATC CTGGCCTCCC
501 AGAGTATTCC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The following *C.pneumoniae* protein (PID 4377373) was also expressed <SEQ ID 305; cp7373>:

```

15 1  MSTTTVKHFI HTASRWEPLV KEIVASNYWH AQWINTLSFL ENSGAKKISA
51 SEHPTVEKKE VLKHAAEEFR HGHYLTQIS RISETSLPDY TSKNLLGGLL
101 TKYYLHLLDL RTCRVLENEY SLSGQTLKTA AYILVTYAE LRASELYPLY
151 HDILKEAQSK ITVKSIILEE QGHLQEMERE LKDLPHGEEL LGYACQFEGE
201 LCLQFVERLE QMIFDPSSTF TKF*

```

20 The cp7373 nucleotide sequence <SEQ ID 306> is:

```

1  ATGTCTACAA CCACAGTAAA ACACCTTTATC CACACAGCCT CTCGTTGGGA
51 GCCCGTTCTC AAAGAGATCG TAGCTTCCAA CTATTGGCAT GCACAATGGA
101 TAAATACCCT GTCCTTTTTA GAAAATAGTG GAGCAAAAAA AATCTCCGCA
151 AGTGAACATC CTACGGAGGT AAAGGAAGAA GTTTTAAAC ATGCTGCTGA
25 201 AGAATTTTCT CATGGTCACT ATCTAAAAAC TCAGATTCTC AGAATCTCAG
251 AGACTTCTCT CCCTGACTAT ACATCTAAAA ATCTTCTGGG AGGCTTACTT
301 ACAAAATATT ACCTCCATCT TCTAGATTTA AGGACGTGCC GAGTACTGGA
351 AAATGAATAC TCCCTATCGG GACAAACGTT AAAAAGTGCA GCGTATATTT
401 TAGTTACCTA CGCAATCGAA CTTCGTGCTT CTGAACCTTA TCCTCTGTAT
30 451 CACGATATTC TGAAAGAAGC TCAAAGTAAA ATAACGGTAA AATCCATTAT
501 CTTAGAAGAG CAAGGCCATC TGCAAGAGAT GGAACGTGAA CTTAAAGATC
551 TCCCCCACGG GGAGGAACCT TTAGGCTATG CTTGCCAATT CGAAGGGGAG
601 CTTTGCTTGC AGTTTGTAGA GAGATTAGAA CAAATGATCT TCGATCCTTC
651 CTCGACTTTT ACAAAGTTCT AG

```

35 The PSORT algorithm predicts cytoplasm (0.1069).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 152A; 6282 = lanes 8 & 9; 7373 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 152B & 153) and for FACS analysis.

40 These experiments show that cp6282 & cp7373 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 154 ,
 Example 155 ,
 Example 156 ,
 Example 157 and
 45 Example 158

The following *C.pneumoniae* protein (PID 4376412) was expressed <SEQ ID 307; cp6412>:

```

1  MSSSEVVFTQ VHGLGFGGLS SKSVVPFKKS LSDAPRVVCS ILVLTGLGLA
51 LVCGIAITCW CVPGVILMGG ICAIVLGAIS LALSFLWLWG LFSNCCGSKR
101 VLPGEGLLRD KLLDGGFSRA APSGMGLPGD GSPRASTPSC LEBLQAEIQ
50 151 VTQAIDQMSD D*

```

The cp6412 nucleotide sequence <SEQ ID 308> is:

-171-

1 ATGAGCAGTT CGGAAGTTGT TTTCCAGACA GTTCATGGCC TTGGCTTTGG
 51 TGGATTGTCT TCAAAAAGTG TTGTCCCTTT TAAGAAAAGT CTTTCGGATG
 101 CGCCCCGTGT TGTGTGCTCG ATTTTAGTTT TGACTCTGGG GTTGGGAGCG
 151 CTTGTTTGTG GTATTGCCAT TACTTGTGGG TGTGTCCCGG GAGTTATTTT
 201 AATGGGGGGA ATTTGCGCTA TAGTTT TAGG TGCAATTTCT TTAGCTTTAA
 251 GTCATATTTG GTTGTGGGGT TTATTTTCTA ATTGTTGTGG TTCTAAGAGA
 301 GTTTTACCGG GTGAGGGGAT GCTACGGGAT AAGCTTTTAG ATGGTGGATT
 351 TTCAAGAGCG GCACCTTCAG GAATGGGACT TCCGGGTGAT GGATCTCCAA
 401 GAGCGTCAAC GCCATCTTGC CTAGAGGAAC TTCAAGCAGA GATACAGGCA
 451 GTTACTCAAG CTATCGATCA GATGTCAGAT GATTGA

The PSORT algorithm predicts inner membrane (0.4864).

The following *C.pneumoniae* protein (PID 4376431) was also expressed <SEQ ID 309; cp6431>:

1 LRAGGSLVTT YPKQGRLRS PEQLRVLDL VQSYPNHLHA IELDCGAIPQ
 51 DLIGATYIIT FADFSTYILS LRSYQANSFS DDTWGIWFGS IDDPVQAVIS
 101 FLKDHGFALP STLQDPLLC TNK*

The cp6431 nucleotide sequence <SEQ ID 310> is:

1 TTGCGAGCAG GAGGTAGTCT TGTTACAACA TACCCTAAGG AAGGTCAGAG
 51 ATTGCGCTCC CCAGAACAGT TAAGAGTTCT GGATGATTA GTGCAAAGCT
 101 ATCCAAATCA CCTACATGCG ATTGAACCTG ATTGTGGTGC AATCCCTCAA
 151 GATTTGATCG GAGCCACCTA TATCATCACG TTCGCCGATT TTCCACCTA
 201 TATTCTCTCT TTAAGAAGCT ACCAAGCCAA TTCTCCCTCC GATGATACAT
 251 GGGGGATTGT GTTTGGATCT ATTGACGATC CTGTTCAAGC AGTCATATCA
 301 TTTTAAAAG ATCATGGATT TGCTCTTCCC TCGACCTTAG CTCAAGATCC
 351 TTTGCTTGT ACTAACAAGT AA

25 The PSORT algorithm predicts cytoplasm (0.2115).

The following *C.pneumoniae* protein (PID 4376443) was also expressed <SEQ ID 311; cp6443>:

1 MIMTTISNSP SPALNPELSL IPPPTLVSSG TQTSLAYTIP AQGRRSTLRI
 51 ILDIFIILG LATIISTFIV IFFLGNLNL STPSIISSSC LIIVGLLFLI
 101 MGLYFMISSL DQGLVGLLQK ELSQAEREE EYIQEIALR GAPRAESFTE
 151 SPSTWL*

The cp6443 nucleotide sequence <SEQ ID 312> is:

1 ATGATTATGA CTACTATATC TAACTCACCC TCCCCTGCAT TGAATCCCCA
 51 ACTTTCCCTT ATTCCTCCAC CAACACTTGT ATCTTCAGGT ACGCAAACAT
 101 CTCTAGCTTA TACGATCCCC GCACAAGGAC GAAGATCCAC CCTACGTATT
 151 ATATTAGATA TATTCAATTAT CATTCCTGGT TTAGCTACGA TCATTCTTAC
 201 CTTTATGTG ATTTCTTTT TAAATGGGCT GAACTTGCTC TCGACCCCAT
 251 CTATTATCTC TTCGTCATGT TTAATCATTTG TTGGATTGCT TTTTGTGATT
 301 ATGGGGTTAT ATTTTCATGAT CTCGAGTTTG GATCAGGGGC TTGTAGGCCCT
 351 TCTGCAAAAG GAACTCTCTC AAGCCGAAGA AAGAGAAGAA GAGTATATCC
 401 AGGAAATCGA AGCTTTAAGA GGAGCTCCTA GAGCAGAATC TCCCACAGAG
 451 TCTCCTAGTA CCTGGTTATG A

The PSORT algorithm predicts inner membrane (0.5585).

The following *C.pneumoniae* protein (PID 4376496) was also expressed <SEQ ID 313; cp6496>:

1 MLIGRYSSDD QFTEATKNTF TIILKGFVRD NLEGLTNPIS EIVSETSSSI
 51 KDSVLRSLPI LGSILGCARL YSTLSTNDPL DETQEKIWHT IFGALETGLG
 101 GILILLEFKII FVILHCIFHL VIGFCK*

The cp6496 nucleotide sequence <SEQ ID 314> is:

1 ATGCTAATAG GCAGATACAG TAGTGATGAC CAATTCACCTG AAGCAACAAA
 51 AAACACCCCA ACCATAATTA AGCTAGGTTT TGTTAGAGAT AATCTCGAGG
 101 GATTAACGAA CCCTATCTCT GAAATCGTCT CGGAAACCTC CTCTTCTATT
 151 AAAGATTCGG TTCTTCGCTC TCTTCCTATT TTAGGGTCCA TTTTAGGATG
 201 CGCCCGACTT TACAGCACAC TCTCTACAAA TGATCCTCTT GACGAAACTC
 251 AAGAAAAGAT TTGGCACACT ATATTGGAG CCTTAGAAAC CTTAGGCTTA
 301 GGGATTCTCA TCCTCTTATT TAAATTTATT TTTGTTATAT TACACTGCAT
 351 ATTTCATCTA GTTATTGGGT TCTGCAATA A

The PSORT algorithm predicts inner membrane (0.5989).

The following *C.pneumoniae* protein (PID 4376654) was also expressed <SEQ ID 315; cp6654>:

```

1  MKTKMNSRKK AGQWAFNSP TPGVSSTLVL AWTWPWGYDYK DVQDILERKD
51 PMSSSLSEKD SKEFLKNLFV DLLENGFTSV HIIHAEFAFTP LDHTGKPHFK
101 RDNVYLPGLK LGALNEAAVQ ANVSADTQFT LFLTQDECNP FHDKKRG*

```

The cp6654 nucleotide sequence <SEQ ID 316> is:

```

1  ATGAAACTA AAATGAAGTC TAGAAAAAAA GCAGGTCAAT GGGCAATTTT
51 CAATTCTCCA ACTCCTGGTG TCAGTTCAAC TTTAGTTTTA GCATGGACTC
101 CTGGGGTTA TTACGACAAG GATGTACAAG ATATCTTAGA AAGAAAAGAT
151 CCGATGAGCT CTTCGCTTTC TGAAAAAGAC TCAAAGGAGT TCTTGAAAAA
201 TCTGTTTGTG GATCTCTTAG AAAATGGCTT CACATCAGTA CATATTACAG
251 CAGAAGAAGC TTTCACCTCT CTTGATCATA CCGGGAAACC TCACTTTAAA
301 AGAGACAATG TGTACTTACC CGGAAAGTTG TTAGCGCCTT TGAATGAGGC
351 TCGCGTACAA GCCAATGTAA GTGCGGATAC TCAATTACAA TTGTTCTTAA
401 CTCAAGATGA GTGCAATCCT TTTCATGATA AGAAAAGAGG TTAA

```

The PSORT algorithm predicts cytoplasm (0.0730).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 154A; 6412 = lanes 2-3; 6431 = lanes 11-12; 6443 = lanes 5-6; 6496 = lanes 8-9; 6654 = lane 10; markers in lanes 1, 4, 7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 154B, 155, 156, 157 & 158) and for FACS analysis.

These experiments show that cp6412, cp6431, cp6443, cp6496 & cp6654 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from their sequences alone.

Example 159 and

Example 160

The following *C.pneumoniae* protein (PID 4376477) was expressed <SEQ ID 317; cp6477>:

```

1  LLKFFLVCEE LCILTVATHR ALLETPLALS FFKELKTKYV YRAKDILQLH
51 NYKGFITLNT SPLCS*

```

The cp6477 nucleotide sequence <SEQ ID 318> is:

```

1  TTGCTAAAGT TCTTCTAGT ATGTGAAGAG TTATGTATAC TTAAGTTTGC
51 TACACATAGA GCTCTCTTAG AAACCTCCTT AGCTCTATCA TTTTAAAG
101 AACTTAAGAC AAAATATGTC TACAGGGCGA AAGACATACT ACAACTACAT
151 AACTATAAAG GATTACTAT CCTTAATACA TCACCGTTAT GTTCTTAA

```

The PSORT algorithm predicts inner membrane (0.128).

The following *C.pneumoniae* protein (PID 4376435) was also expressed <SEQ ID 319; cp6435>:

```

1  LWSHFPRGFF MLPFCPTILL AKPFLNSEN YGLERLAATVD SYFDLQGSQI
51 VFSLKQDQGI TVEELSAKDR KFKPGSMNCT LYTEDPILPA HNSFNSCSDI
101 QMRTPISPIH *

```

The cp6435 nucleotide sequence <SEQ ID 320> is:

```

1  TTGTGGTCGC ATTTCCCAAG AGGATTTTTT ATGCTCCCTT TTTGCCCTAC
51 CATCCTTCTT GCTAAACCTT TTTTAAATAG CGAGAATTAC GGCTTAGAAC
101 GTTTAGCTGC AACCCTAGAT TCTTATTTTG ATCTGGGACA GTCTCAAATA
151 GTCTTCCTAA GCAAACAGGA TCAAGGAATC ACTGTGGAAG AATTGAGTGC
201 TAAAGATAGG AAATTCAGC CAGGCTCTAT GAACTGTACA CTGTACACTG
251 AAGATCCTAT CTTACCTGCT CATAATTCCT TTAGTAATTG CTCTGATATT
301 CAAATGCGTA CTCCGATTAG CCCTATACAT TAA

```

The PSORT algorithm predicts periplasmic space (0.4044).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 159A; 6435 = lanes 2-4; 6477 = lanes 5-7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 159B & 160) and for FACS analysis.

- 5 These experiments show that cp6477 & cp6435 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequences alone.

**Example 161 and
Example 162 and
Example 163**

- 10 The following *C.pneumoniae* protein (PID 4376441) was expressed <SEQ ID 321; cp6441>:

```

1  VEAGANVLVI DTAHAHSGKV FQTVLEIKSQ FPQISLVVGN LVTAEAAVSL
51  AEIGVDAVKV GIGPGSICTT RIVSGVGYPQ ITAITNVAKA LKNSAVTVIA
101 DGRIRYSGDV VKALAAGADC VMLGSLLAGT DEAPGDIVSI DEKLFKRYRG
151 MGSLGAMKQG SADRYFQTQG QKKLVPGGVE GLVAYKGSVH DVLYQILGGI
15  201 RSGMGYVGAE TLKDLKTKAS FVRITESGRA ESHIHNIYKV QPTLNY

```

The cp6441 nucleotide sequence <SEQ ID 322> is:

```

1  GTGGAAGCTG GAGCAAATGT TCTAGTCATT GACACAGCTC ATGCACACTC
51  TAAAGGAGTA TTCCAAACAG TTTTAGAAAT AAAATCCCAG TTCCACACAA
20  101 TTTCTTTAGT TGTAGGGAAT CTTGTTACAG CTGAAGCCGC AGTTTCCTTA
151 GCTGAGATTG GAGTTGACGC TGTAAAGGTA GGTATTGGCC CAGGATCTAT
201 CTGTACAAC TGAATCGTTT CAGGGGTCGG TTATCCACAA ATTACTGCCA
251 TTACAAACGT AGCAAAAGCT CTTAAAAACT CTGCCGTGAC TGTAAATGCT
301 GATGGGAGAA TCCCGTATTC TGGAGATGTG GTAAAAGCAT TAGCAGCAGG
351 AGCAGACTGT GTCATGCTAG GAAGTTTGCT TGCAGGACT GATGAAGCTC
25  401 CTGGGGATAT CGTTTCTATC GATGAGAAGC TTTTAAAAG GTACCGCGGC
451 ATGGGATCTT TAGGCGCTAT GAAACAAGGA AGTGCTGACC GGTATTTTCA
501 AACACAGGGA CAGAAAAAGC TGGTTCCTGG GGGAGTTGAA GGAAGTAGTCG
551 CTTATAAAGG CTCTGTCCAC GATGTCCTCT ATCAAAATTT AGGAGGAATA
601 CGCTCAGGTA TGGGGTATGT TGGAGCTGAA ACTCTCAAAG ATTTAAAAAC
30  651 TAAGGCTTCC TTTGTTTCGAA TTAAGTGAATC TGAAGAGCT GAAAGTCATA
701 TTCATAATAT TTACAAAGTT CAACCAACCT TAAATTATTA A

```

The PSORT algorithm predicts bacterial inner membrane (0.132).

The following *C.pneumoniae* protein (PID 4376748) was also expressed <SEQ ID 323; cp6748>:

```

1  LFSEGTALNL FRIFAPLRNR VTTEYSRRARQ PDLHRIAIVY IGVLDSESSK
35  51  ILERLISYMS CIYSESQMYL RFFMGKNVNO SAVLSKLHVE NLHIRCGFFS
101 EDAVPESEPF DLSIYVHTDR SCPLPTKRRS SSWELQTVEL PESIYPQSEF
151 LLMRPRMLS*

```

The cp6748 nucleotide sequence <SEQ ID 324> is:

```

1  TTGTTCTCTG AGGGGACAGC TCTAAATTTA TTTCGTATAT TTGCTCCACT
40  51  ACGCAACCGT GTGACTACAG AATACAGTCG TGCTAGGCCA CCCGACCTAC
101 ATAGAATTGC CATCGTCTAT ATAGGAGTTC TCGATTGAGA AAGTTCCAAG
151 ATCCTAGAGC GGCTAATCTC TTATATGAGT TGTATCTATT CTGAATCGCA
201 AATGTATTTA AGATTCTTTA TGGGCAAGAA TGTAATACAA AGTGCTGTAC
251 TCTCAAAATT ACATGTAGAA AATCTGCACA TCCGTGTGGG GTTTTTCAGC
45  301 GAGGATGCTG TTCCAGAGAG TGAGCCCTTC GATCTCTCCA TCTACGTGCA
351 CACAGATCGT AGCTGTCTCT TCCCTACGAA AAAACGGAGC AGCTCCTGGG
401 AACTCCAAC TGTAGAAGCT CCAGAGTCAA TATATCCACA GTCGGAATTC
451 CTATTGATGA GACCTCGAAT GCTTTCGTAG

```

The PSORT algorithm predicts cytoplasm (0.170).

- 50 The following *C.pneumoniae* protein (PID 4376881) was also expressed <SEQ ID 325; cp6881>:

-174-

1 MRPHRKHVSS KSLALKQSAS THVEITTKAF RLSMPLKQLI LEKSDHLPPM
 51 ETIRVVLTSK KDKLGTEVHV VASHGKEILQ TKVHNANPYT AVINAFKKIR
 101 TMANKHSNKR KDRTKHDGLL AAKEERIAIQ EEQEDRLSNE WLPVEGLDAW
 151 DSLKTLGYVP ASAKKKISKK KMSIRMLSQD EAIRQLESAA ENFLIFLNEQ
 201 EHKIQCIFYK HDGNYVLIEP SLKPGFCI*

The cp6881 nucleotide sequence <SEQ ID 326> is:

1 ATGAGACCTC ATCGTAAACA CGTATCATCT AAAAGCTTAG CTTTAAAGCA
 51 ATCTGCATCA ACTCATGTAG AGATCACAAC AAAAGCCTTT CGTCTCTCTA
 101 TGCCTCTAAA ACAGCTGATC CTAGAGAAAA GCGACCACCT CCCCCCTATG
 151 GAAACAATCC GTGTGGTGCT AACCTCTCAT AAAGATAAGC TAGGCACCGA
 201 GGTGCATGTT GTAGCTTCTC ATGGCAAAGA AATCCTTCAA ACTAAGGTTT
 251 ATAACGCAA CCCATACACT GCAGTGATCA ATGCTTTTAA GAAAAATCCG
 301 ACCATGGCAA ATAAGCACTC CAATAAACGT AAAGACAGGA CAAAACATGA
 351 TCTAGGTCTT GCAGCAAAG AAGAACGTAT CGCAATACAG GAAGAACAAG
 401 AAGATCGCCT TAGCAACGAG TGGCTTCCTG TCGAAGGCCT CGATGCCTGG
 451 GATTCTCTAA AAACCTTGG GTATGTCTCC GCATCAGCGA AAAAGAAGAT
 501 CTCCAAGAAA AAGATGAGCA TTCGTATGCT ATCTCAAGAC GAGGCTATCC
 551 GCCAGCTAGA GTCTGCCGCA GAAAACTTCC TGATCTTCTT GAACGAGCAA
 601 GAGCATAAAA TCCAATGCAT TTATAAAAAA CATGACGGCA ACTATGTCCT
 651 TATTGAACCT TCCCTCAAGC CAGGATTCTG CATCTGA

The PSORT algorithm predicts cytoplasm (0.249).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 161A; 6441= lanes 7-9; 6748 = lanes 2-3; 6881 = lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 161B, 162 & 163) and for FACS analysis.

25 These experiments show that cp6441, cp6748 & cp6881 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 164 and Example 165 30 Example 166

The following *C.pneumoniae* protein (PID 4376444) was expressed <SEQ ID 327; cp6444>:

1 MEQPNCVIQD TTTVLYALNS FDPRLSDDTH RLKQSPLEA ENALGEFIEG
 51 LDNNSFPLEE VAIPILPGYH PKFYLSFIDR DDQGVHYEVL DGVFLKTVAA
 101 CIIENSFLTD SMSPELLSEV KEALKR*

35 The cp6444 nucleotide sequence <SEQ ID 328> is:

1 ATGGAGCAAC CCAATTGTGT GATTCAAGGAT ACTACAACCTG TTTTGTATGC
 51 CTAAATAGC TTTGATCCTA GACTTAGTGA TGACACTCAC AGACTTGGGA
 101 AGCAATCACC TCTTGAAGCA GAAAATGCTC TTGGAGAATT TATTGAAGGT
 151 TTGGATACAA ATAGCTTTCC TTTAGAGGAA GTTGCCATTC CCATCCTGCC
 40 201 AGGTTATCAC CCTAAGTTT ATTTATCTTT CATAGATAGG GACGATCAG
 251 GTGTCCACTA TGAAGTTTGA GATGGCGTAT TTTTAAAGAC AGTCGCTGCT
 301 TGTATTATAG AGAACTCCTT CTTAAGTATG TCTATGAGCC CGGAGCTTCT
 351 CAGCGAAGTT AAGGAAGCTC TGAACGATG A

The PSORT algorithm predicts cytoplasm (0.2031).

45 The following *C.pneumoniae* protein (PID 4376413) was also expressed <SEQ ID 329; cp6413>:

1 MAVQSIKEAV TSAATSVGCV NCSREAIAP NTEERATSIA RSVIAAIIAV
 51 VAISLLGLGL VVLAGCCPLG MAAGAITMLL GVALLAWAIL ITLRLLNIFK
 101 AEIPSPGNNG EPNERNSATP PLEGGVAGEA GRGGGSPLTQ LDLNSGAGS*

The cp6413 nucleotide sequence <SEQ ID 330> is:

50 1 ATGGCTGTTT AATCTATAAA AGAAGCCGTA ACATCAGCCG CAACATCAGT

51 AGGATGTGTA AACTGTTCTA GAGAGGCTAT ACCAGCATTT AATACAGAGG
 101 AGAGAGCAAC GAGTATTGCT AGATCTGTTA TAGCAGCTAT CATTGCTGTT
 151 GTAGCTATCT CCTTACTCGG ACTAGGTCTT GTAGTCTTTG CTGGTTGCTG
 201 TCCTTTAGGA ATGGCTGCGG GTGCTATAAC AATGCTGCTG GGTGTAGCAT
 251 TATTAGCTTG GGCAATACTG ATTACTTTGA GACTGCTTAA TATACCTAAG
 301 GCTGAAATAC CGAGTCCAGG GAACAACGGT GAGCCTAATG AAAGAAATTC
 351 AGCAACTCCT CCTCTAGAGG GTGGTGTGTC AGGAGAAGCC GGTCCGCGCG
 401 GGGGGTCACC TTTAACCCTA CTTGATCTCA ATTGAGGGGC GGGAGTTAG

The PSORT algorithm predicts inner membrane (0.6180).

10 The following *C.pneumoniae* protein (PID 4377391) was also expressed <SEQ ID 331; cp7391>:

1 MMLRVIELPL LPIKQALEKA FVQYNSYKAK LTKVEPCFRE SPAYITSEER
 51 LQSLDQTLER AYKEYQKRFO EPSRLESEVS GCREHLREQV KQFETQGLDL
 101 IKEELIFVSD VLFRKMVSCL VSTVHVPFME FYEYFELHR LRLRAQWMAN
 151 AEIYSKVRKA FPEMLKETLE KAKAPREEEY WLLCEERKSK EKRLILNKIE
 201 AAQQRVKDLE PPPIKETGKQ KKKKEYSFFI RLKS*

The cp7391 nucleotide sequence <SEQ ID 332> is:

1 ATGATGCTTC GTGTCATAGA GCTTCCACTA CTTCTATATA AGCAAGCGTT
 51 GGAGAAGGCT TTTGTACAAT ATAATAGCTA CAAAGCGAAG TTAACCAAGG
 101 TAGAACCTTG CTTTAGAGAG AGCCCTGCCT ATATAACTAG CGAAGAGCGA
 151 CTCCAGAGTT TGGATCAGAC TTTAGAACGT GCGTACAAAG AGTACCAGAA
 201 GAGATTCCAG GAGCCTTCAC GTTTGGAATC GGAAGTAAGT GGATGTAGAG
 251 AGCATCTTAG AGAGCAGGTA AAACAATTG AACTCAAGG ACTAGACTTG
 301 ATCAAAGAAG AGCTTATTTT TGTTAGTGAT GTGTTATTCC GAAAAATGGT
 351 CAGTGTCTTA GTGTCGACAG TGCATGTTCC CTTTATGGAG TTTTATTATG
 401 AGTATTTTGA GTTGATAGA TTGAGGTTGC GGGCCCAATG GATGGCGAAT
 451 GCCGAGATTT ATAGCAAAGT TAGAAAAGCA TTCCAGAGA TGTGAAGGA
 501 GACCTTAGAA AAAGCTAAG CTCCAGAGA AGAAGAGTAT TGTTACTTTT
 551 GCGAGGAGAG AAAGAGTAAG GAGAAGCGTT TGATTCTCAA CAAGATAGAG
 601 GCAGCTCAGC AGCGGGTAAA AGATTAGAA CCTCCTCCTA TTAAAGAGAC
 651 AGGGAACAG AAACGGAAGA AAGAATATTC GTTTTTCATT CGATTAAAT
 701 CGTGA

The PSORT algorithm predicts inner membrane (0.1489).

The proteins were expressed in *E.coli* and purified as his-tag and GST-fusion products (Figure 164A; 6444=lanes 11-12; 7391=lanes 2-3; 6413=lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 164B, 165 & 166) and for FACS analysis.

These experiments show that cp6444, cp6413 & cp7391 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 167 ,
 Example 168 ,
 Example 169 and
 Example 170

The following *C.pneumoniae* protein (PID 4376463) was expressed <SEQ ID 333; cp6463>:

1 MKKKVTIDEA LKEILRLEGA ATQEELCAKL LAQGFATTQS SVSRWLRKIQ
 51 AVKVAGERGA RYSLPSTTEK TTTRHLVLSI RHNASLIVIR TVPGSASWIA
 101 ALLDQGLKDE ILGTLAGDDT IFVTPIDEGR LPLLMVSIAN LLQVFELD*

The cp6463 nucleotide sequence <SEQ ID 334> is:

1 ATGAAAAAAA AAGTAACAT AGATGAGGCT TAAAAAGAAA TTTTACGTCT
 51 TGAAGGAGCG GCAACTCAGG AGGAATTATG TGCAAACTC TTAGCTCAAG
 101 GTTTTGCTAC AACCCAGTCG TCTGTATCTC GTTGGCTACG AAAGATTACG
 151 GCTGTAAAGG TTGCTGGAGA GCGTGGTGCT CGTTATTCTT TACCTCTTC

201 AACAGAGAAG ACCACGACCC GTCATTGGGT GCTCTCTATT CGCCATAACG
 251 CCTCTCTTAT TGTAATTCGT ACGGTTCCCTG GTTCAGCTTC TTGGATCGCT
 301 GCTTTGTTAG ATCAAGGGCT CAAAGATGAA ATTCTTGGA CTTTGGCAGG
 351 AGATGACACG ATTTTGTGCA CTCCTATAGA TGAAGGGAGG CTCCCATTTG
 401 TGATGGTTTC GATTGCAAAT TTAAGTCAAG TTTTCTTGA TTA

The PSORT algorithm predicts inner membrane (0.1510).

The following *C.pneumoniae* protein (PID 4376540) was also expressed <SEQ ID 335; cp6540>:

1 MSQCQSSSTS TWEWMKSFVP NWKNPTPLPS PIPSEDEFIL AYEPFVLPKT
 51 DPENANQANPP GTSTPNVENG IDDLNPLLGQ PNEQNANNP GTSGSNPTSL
 101 PAPERLPETE ENSQEEBQGS QNNEDLIG*

The cp6540 nucleotide sequence <SEQ ID 336> is:

1 ATGTCTCAAT GTCAGAGTAG CAGTACATCT ACCTGGGAAT GGATGAAATC
 51 TTTTGTGCCA AACTGGAAGA ATCCAACCTCC CCCCTTATCT CCTATACCTT
 101 CTGAGGACGA ATTTATATTA GCATACGAGC CATTGTGTCT ACCGAAAACA
 151 GATCCAGAAA ACGCACAGC TAATCCTCCA GGCACATCTA CACCGAATGT
 201 AGAAAACGGG ATCGATGATC TCAACCTCT TCTGGGGCAA CCCAACGAAC
 251 AAAACAATGC CAACAATCCA GGAACCTCTG GATCTAATCC TACATCTCTA
 301 CCCGCCCCCG AACGACTCCC TGAAACTGAA GAGAACAGCC AAGAAGAAGA
 351 ACAAGGATCT CAAATAATG AGGATCTTAT AGGATAA

20 The PSORT algorithm predicts cytoplasm (0.3086).

The following *C.pneumoniae* protein (PID 4376743) was also expressed <SEQ ID 337; cp6743>:

1 LREEGSVSFR EYFRAYMCDK IVAQKNFLFT LDAVIKQAGW RSQEKLNLFY
 51 VESQALGREI KVSLEBYIQS MVGILGSQRT KKSFKFSVDF TPLEQALQER
 101 CSSDDDEDAT ATSTATGATA SPTDMHEDE*

25 The cp6743 nucleotide sequence <SEQ ID 338> is:

1 TTGAGAGAAG AAGGTAGTGT TTCTTTCAGA GAATATTTCA GAGCCTATAT
 51 GTGTGATAAA ATCGTGGCAC AGAAGAACTT CTTATTTACT TTAGACGCTG
 101 TAATTAAACA GGCCGGTTGG AGATCACAAG AGAAACTCAA TTTATTTTAT
 151 GTTGAAAGTC AGGCTTTAGG AAGAGAAATC AAAGTCAGCT TAGAGGAATA
 201 TATTCAGAGT ATGGTCGGGA TTTTGGGATC TCAGAGAACC AAGAAAAGCT
 251 TTAAGTTTTC TGTCGACTTT ACCCCTTTAG AGCAGGCTCT ACAGAAAGA
 301 TGCTCTTCTG ATGATGACGA AGATGCAACA GCAACTTCGA CCGCTACAGG
 351 GGCAACAGCA TCTCCGACTG ACATGCACGA AGATGAGTAA

The PSORT algorithm predicts cytoplasm (0.2769).

35 The following *C.pneumoniae* protein (PID 4377041) was also expressed <SEQ ID 339; cp7041>:

1 MLMMLMIIG ITGGSGAGKT TLTQNIKEIF GEDVSVICQD NYVKDRSHYT
 51 PEERANLIWD HPDAFDNDLL ISDIKRLKNN EIVQAPVDF VLGNRSKTEI
 101 ETIYPSKVL VEGILVFENQ ELRDLMDIRI FVDTDADERI LRRMVRDVQE
 151 QGDSVDCIMS RYLSMVKPMH EKFIETPKY ADIIVHGNRY QNVVTNLSQ
 201 KIKNHLENAL ESDETYVMVN SK*

The cp7041 nucleotide sequence <SEQ ID 340> is:

1 ATGTTGATGA TGCTTATGAT GATTATTGGA ATTACAGGAG GTTCTGGAGC
 51 TGGGAAAACC ACCCTAACCC AAAACATTAA AGAAATTTTC GGTGAGGATG
 101 TGAGTGTTAT CTGCCAAGAT AATTATTACA AAGATAGATC TCATTATACT
 151 CCTGAAGAAC GTGCCAATT AATTGGGAT CATCCGGACG CCTTTGATAA
 201 TGAATTTATTA ATTTTCAGACA TAAACGCTCT AAAAAATAAT GAGATTGTCC
 251 AAGCCCCAGT TTTTGAATTT GTTTAGGTA ATCGATCTAA AACGGAGATA
 301 GAAACGATCT ATCCATCTAA AGTTATTCTT GTTGAAGGTA TTCTGGTCTT
 351 TGAAAAATCAA GAACTTAGAG ATCTTATGGA TATTAGGATC TTGTAGACA
 401 CCGATGCTGA TGAAAGGATA CTACGCCGTA TGGTTCGAGA TGTTCAGGAA
 451 CAAGGAGATA GCGTGGACTG CATCATGTCT CGTTATCTTT CTATGGTAAA
 501 GCCTATGCAT GAGAAATTTA TAGAGCCGAC TCGGAAATAT GCTGATATCA
 551 TTGTACATGG AAATTACCGA CAAAACGTAG TAACAAATAT TTTGTACAG
 601 AAAATTAAAA ATCATTTAGA GAATGCCCTG GAAAGCGATG AGACGTATTA
 651 TATGGTCAAC TCTAAGTAA

The PSORT algorithm predicts inner membrane (0.1022).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 167A; 6463 = lanes 2-4; 6540 = lanes 5-7; 6743 = lanes 8-9; 7041 = lanes 10-11). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 167B, 168, 169 & 170) and for FACS analysis.

These experiments show that cp6463, cp6540, cp6743 & cp7041 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 171 and Example 172 and Example 173

The following *C.pneumoniae* protein (PID 4376632) was expressed <SEQ ID 341; cp6632>:

```

1  VQLFQYMNES GWDWLCDFDS QGEGFQLSRL VGLLHSSWAL YEAKEQFYLP
51 EVSLLTWEEEL IEMQLLSKPT KHGVAKDLN VFEKHFQRF R QYLGSLDLNQL
101 RFENTFLNYP KYHLDR*
```

The cp6632 nucleotide sequence <SEQ ID 342> is:

```

1  GTGCAATTAT TTCAATATAT GAATGAGTCC GGATGGGATT GGCTTTGTGA
51 TTTTGATTCT CAAGGCGAGG GATTCCAGTT ATCACGTCTG GTTGGGCTGT
101 TACATTCTGTC CTGGGCATTA TACGAAGCAA AAGAGCAATT TTACCTTCCT
151 GAGGTTCTCT TATTGACCTG GGAAGAACTG ATAGAAATGC AGTTATTAAG
201 CAAACCAACA AAACACGGGG TTGCAAAAGA TCTTTGTAAT GTATTTGAAA
251 AACACTTTCA AAGGTTTAGA CAGTACCTAG GTTCCTTAGA TCTAAATCAA
301 AGGTTCCGAA ATACCTTCTT GAATTATCCT AAATACCATT TAGATAGGGA
351 GTGA
```

The PSORT algorithm predicts cytoplasm (0.3627).

The following *C.pneumoniae* protein (PID 4376648) was also expressed <SEQ ID 343; cp6648>:

```

1  MPVSSAPLPT SHRPSSGNLG LMEPNKALK AKHQDKTKT IKLLVKILVA
51 ILVIEVLGII AAFPIGTPP ICLILGGLI LTTVLCVLLL VIKLALVNKT
101 EGTAEQQIK RKLSSKSSIS*
```

The cp6648 nucleotide sequence <SEQ ID 344> is:

```

1  ATGCCCGTGT CCTCAGCCCC CCTACCCACA AGCCACCGCC CTTCTCTGG
51 AAATCTAGGC CTCATGGAAC CAAATTCCTAA AGCTCTAAA GCAAAGCATC
101 AAGATAAAAC GACGAAGACG ATTAAACTTT TAGTTAAAAT CCTTGTGGC
151 ATTCTAGTAA TAGAAGTTTT AGGAATAATT GCAGCTTTCT TTATTCCTGG
201 GACTCCTCCC ATCTGCTTGA TTATCCTAGG AGGCCTTAT CTTACAACAG
251 TACTCTGTGT GCTTCTTCTT GTTATAAAGC TTGCCCTTGT AAACAAAACC
301 GAAGGAACAA CTGCTGAACA GCAGATAAAA CGTAAACTCT CTTCTAAAAG
351 TATTTCTTAG
```

The PSORT algorithm predicts inner membrane (0.6074).

The following *C.pneumoniae* protein (PID 4376497) was also expressed <SEQ ID 345; cp6497>:

```

1  MKPNSIIFLE NTKHYPDIFR EGFVRDRHGL MEASDWLLST EITIIRSILG
51 AIPILGNILG AGRLYSVWYT SDEWKKQVV *
```

The cp6497 nucleotide sequence <SEQ ID 346> is:

```

1  ATGAAGCCAA ATAGTATTAT TTTTITAGAA AATACTAAGC ATTATCCCGA
51 CATCTTTCGA GAAGGATTTC TTCGTGATCG TCATGGACTA ATGGAAGCCT
101 CGGATTGGTT ACTTCTACG GAAATTACGA TCATTGCTC CATTTCTGGG
151 GCTATCCCTA TTTTAGGAAA TATCTTTGGA GCCGACGAC TCTATAGCGT
```

201 TTGGTATACA AGTGACGAAG ATTGGAAAAA ACAAGTGGTT TGA

The PSORT algorithm predicts inner membrane (0.145).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 171A; 6632 = lanes 5-7; 6648 = lanes 8-10; 6497 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 171B, 172, 173) and for FACS analysis.

These experiments show that cp6632, cp6648 and cp6497 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 174 ,
Example 175 ,
Example 176 ,
Example 177 and
Example 178

The following *C.pneumoniae* protein (PID 4377200) was expressed <SEQ ID 347; cp7200>:

1 MPVPIDNSSR NLQEVPESE DLEQHAESP THQSAESSSL QLSLASSAIS
51 SRVEQLSSLV LGMENSDFSS LRDVPIFSAI YESSTHTPVP TPLVGVGYIN
101 GSQSGYYDTQ RESLHLSQLL GSRRVEVVYN QGNFMEASLL NLCPRRPRRD
151 PSPISLALLE LWEAFFLEHP PGSTFNPIFF W*

The cp7200 nucleotide sequence <SEQ ID 348> is:

1 ATGCCCCGTC CTATAGATAA TTCCTCTCGC AACCTACAAG AAGTTCAGAG
51 AAGCCTAGAA GACCTCGAAC AACACGCAGA AGAATCTCCT ACTCATCAAA
101 GTGCAGAAAG CAGTTCCTTG CAACTGTCTC TAGCCTCCTC AGCAATTCTT
151 AGTAGAGTAG AACAACTATC TTCCCTCGTC TTAGGAATGG AAAATTCAGA
201 TTTCTCCTCT TTAAGAGACG TTCCTATCTT CTCAGCTATC TACGAATCTT
251 CAACACACAC ACCTGTCCCC ACTCCTCTAG TTGGCGTGGG ATATATCAAC
301 GGAAGTCAAT CAGGATACTA CGATACACAA AGAGAATCTC TTCACCTCAG
351 CCAATTGTTA GGAAGCCGAA GAGTTGAAGT TGTCTATAAC CAAGGAACT
401 TCATGGAGGC CTCTTTGCTA AATCTGTGCC CCAGAAGACC TCGAAGAGAT
451 CCCTCTCCAA TTTCTTTAGC TCTATTAGAG CTCTGGGAAG CATTTTTTTT
501 AGAACACCCC CCAGGTAGCA CTTTAAATCC AATATTTTTT TGGTAA

The PSORT algorithm predicts cytoplasm (0.3672).

The following *C.pneumoniae* protein (PID 4377235) was also expressed <SEQ ID 349; cp7235>:

1 LNFVSTLTGS DFYAPVLEKL EBAFADTTGQ VILFSSSPDF IVHPIAQQLG
51 ISSWYASCYR DQSAEQTIYK KCLTGDKKAQ ILSYIKKINQ ARSHTFSDHI
101 LDLPFLMLGE EKTVVRPQGR LKKMAKYYW NIV*

The cp7235 nucleotide sequence <SEQ ID 350> is:

1 TTGAATTTTG TATCGACTCT GACCGGCTCC GATTTTATG CTCCTGTTTT
51 AGAAAACTA GAAGAAGCTT TTGCAGATAC CACAGGACAG GTGATCCTTT
101 TTTCTTCTTC TCCAGACTTT ATTGTCCACC CCATAGCGCA GCAACTCGGG
151 ATTAGTTCTT GGTATGCGTC GTGTTATCGC GATCAGTCTG CAGAACAGAC
201 GATCTATAAA AAATGTCTTA CAGGGGATAA AAAAGCGCAA ATTTTGAGTT
251 ATATTAAAAA AATTAATCAA GCAAGAAGCC ATACCTTCTC CGACCATATT
301 TTAGATCTTC CTTTCTTAT GCTGGGAGAA GAGAAAACCG TCGTTCGCCC
351 TCAGGGACGA CTAAGAAAAA TGGCAAAAAA ATATTACTGG AATATCGTTT
401 AA

The PSORT algorithm predicts cytoplasm (0.3214).

The following *C.pneumoniae* protein (PID 4377268) was also expressed <SEQ ID 351; cp7268>:

1 MMHRYFIPLL ALLIFSPSLV RAELOPSENK KGGWPTQLSC AEGSQLFCKF

51 EAAYNNAIEE GKPGILVFFS ERPTPEFADL TNGSFSLSSTP IAKGFNVVVL
 101 CPGLISPLDF FHKMDPVILY MGSFLEMFPE VEAUSGPRLC YILIDEQGGA
 151 QCQAVLPLET KN*

The cp7268 nucleotide sequence <SEQ ID 352> is:

5 1 ATGATGCACC GTTATTTTAT TCCTTTATTA GCACTTCTCA TTTTCTCTCC
 51 TTCTTTAGTC AGGGCAGAGC TACAACCAAG TGAAAACAGA AAAGGGGGGT
 101 GGCTTACACA ACTTTCCTGT GCAGAAGGTT CGCAACTCTT CTGTAAATTC
 151 GAAGCTGCCT ATAATAATGC AATTGAGGAA GGGAAACCTG GGATTTTAGT
 201 CTTTTTCTCT GAGCGACCCA CACCAGAATT TGCCGACTTA ACGAATGGTT
 10 251 CATTTTCTCT CTCTACGCCA ATCGCCAAGG GCTTTAATGT CGTTGTGTGA
 301 TGCCCCGGGC TTATCAGTCC CTTAGACTTT TTCCACAAAA TGGATCCTGT
 351 GATTCTCTAT ATGGGAAGTT TTCTAGAGAT GTCCCTGAA GTGGAGGCAG
 401 TTAGTGGCCC TCGCTTATGT TATATCTTAA TAGATGAACA GGGTGGGGCT
 451 CAATGTCAGG CTGTCCTGCC TTAGAAACA AAGAATTAG

15 The PSORT algorithm predicts inner membrane (0.1235).

The following *C.pneumoniae* protein (PID 4377375) was also expressed <SEQ ID 353; cp7375>:

1 MQRILIVGID TGVGKTIVSA ILARALNAEY WKPIQAGNLE NSDSNIVHEL
 51 SGAYCHPEAY RLHKPLSPHK AAQIDNVISIE ESHICAPKTT SNLIETSGG
 101 FLSPCTSKRL QGDVFSWSC SWILVSQAYL GSINHTCLTV EAMRSRLNI
 20 151 LGMVVNGYPE DEEHWLTQEI KLPIIGTLAK EKEITRTIIS CYAEQWKEVW
 201 TSNHQIGQGV SGTPSLNLH*

The cp7375 nucleotide sequence <SEQ ID 354> is:

1 ATGCAACGTA TCATCATTGT AGGAATCGAC ACTGGCGTAG GAAAAACCAT
 51 TGTCAAGTCT ATCCTTGCTA GAGCACTTAA CGCAGAATAC TGGAAACCTA
 25 101 TACAAGCAGG GAATCTAGAA AATTCAGATA GCAATATTGT TCATGAGCTA
 151 TCGGGAGCCT ACTGTCATCC CGAAGCTTAT CGATTGCATA AGCCCTTGTC
 201 TCCACACAAG GCAGCGCAAA TCGATAATGT AAGTATCGAA GAGAGTCATA
 251 TTTGTGCGCC AAAACAACCT TCGAATCTGA TTATTGAGAC TTCAGGAGGA
 301 TTTTATATCC CCTGCACATC AAAAAGACTT CAGGGAGATG TGTTCCTTC
 30 351 TTGGTCATGT TCTTGATTG TAGTGAGCCA AGCATATCTC GGAAGTATCA
 401 ATCACACCTG TTTAACGGTA GAAGCAATGC GCTCACGAAA CCTCAATATC
 451 TTAGGTATGG TGGTAAATGG GTATCCAGAG GACGAAGAGC ACTGGCTAAC
 501 TCAAGAAATC AAGCTTCCTA TAATCGGGAC TCTTGCCAAG GAAAAAGAAA
 551 TCACAAAGAC AATCATAAGC TGTATGCCC AACAAATGGAA GGAAGTATGG
 35 601 ACAAGCAATC ATCAGGGAAT TCAGGGTGTA TCTGGCACCC CTTACTCTAA
 651 TCTGCATTAG

The PSORT algorithm predicts cytoplasm (0.0049).

The following *C.pneumoniae* protein (PID 4377388) was also expressed <SEQ ID 355; cp7388>:

1 MQVLLSPQLP PPPQHSVSGSI SSPSKLRVLA ITFLVFGMLL LISGALFLTL
 40 51 GIPGLSAAIS FGLGIGLSAL GGVLMISGLL CLLVKREIPT VRPEEIPGV
 101 SLAPSEEPAL QAAQKTLAQL PKELDQLDLD IQEVFACLRLK LKDSKYESRS
 151 FLNDAKKELR VFDFVVEDTL SEIFELRQIV AQEGWDLNFI INGRSLMMT
 201 AESESLDLFH VSKRLGYLPS GDVRGGLKK SAKKIVARLM SLHCBIHKVA
 251 VAFDRNSYAM AEKAFKALG ALEESVYRSL TQSYRDKFLE SERAKIPWNG
 45 301 HITWLRDDAK SGCAEKKLRL AEERWKKFRK AVFWVEEDGG FDINNLLGDW
 351 GTVLDPYRQE RMDEITFHEL YEKTTFLKRL HRKCALAKTT FEKKRSKKNL
 401 QAVEEANARR LKYVRDWDYDQ EFQKAGERLE KLHALYPEVS VSIRENKIQE
 451 TRSNLEKAYE AIEENYRCCV REQEDYWKEE EKREAEFRER GNKILSPEEL
 501 ESSLEQFDHG LKNFSEKLME LEGHLLKLQK EATAEVENKI LSDAESRLEI
 551 VFEDVKEMPC RIEEIEKTLR MAELPLLPTK KAFKACSQY NSCAEMLEKV
 601 KPYCKESLAY VTSKERLVSL DEDLRRAYTE CQKRFQGDG LESEVRACRE
 651 QLRERIQEFQ TQGLDLVEKE LLCVSSRLRN TECDCVSGVK KEAPPKGFY
 701 AQYYDEIYRV RVQSRWMTMS ERLREGVQAC NKMLKAGLSE BDKVLKEEY
 751 WLYREERKNK EKRLVGTKIV ATQQRVAAPF SIEVPEIPEA PEEKPSLLDK
 55 801 ARSLFTREDH T

The cp7388 nucleotide sequence <SEQ ID 356> is:

1 ATGCAAGTAC TTCTATCTCC GCAGCTACCC CCCCCCCCC AACACTCTGT
 51 AGGTCGATT TCTTCTCCAT CTAAACTTCG CGTTTATAGC ATTACTTTTT

5
 10
 15
 20
 25
 30
 35
 40
 45

```

101 TAGTTTTTGG TATGCTCTTA CTGATTTCAG GAGCTCTCTT TCTGACGTTA
151 GGGATTCCAG GATTGAGTGC AGCAATTTCT TTTGGATTAG GCATCGGTCT
201 CTCCGCATTA GGAGGAGTGC TGATGATTTT GGGACTACTA TGCTCTTTAG
251 TAAAACGAGA GATTCCGACA GTACGACCAG AAGAAATTCC TGAAGGGGTT
301 TCGCTGGCTC CTTCTGAGGA GCCAGCTCTA CAGGCAGCTC AGAAGACTTT
351 AGCTCAGCTG CCTAAGGAAT TGGATCAGTT AGATACAGAT ATTCAGGAAG
401 TGTTCCGATG TTTAAGAAAG CTGAAAGATT CTAAGTATGA AAGTCGAAGT
451 TTTTAAACG ATGCTAAGAA GGAGCTTCGA GTTTTGTACT TTGTGGTTGA
501 GGATACCCTC TCGGAGATTT TCGAGTTGCG GCAGATTGTG GCTCAAGAGG
551 GATGGGATTT AAACCTTTTGG ATCAATGGGG GACGAAGCCT CATGATGACT
601 GCAGAATCTG AATCGCTTGA TTTGTTTCAT GTATCGAAGC GGTAGGGTGA
651 TTTACCTTCT GGGGATGTTT GAGGGGAGGG GTTAAAGAAA TCTGCGAAGG
701 AGATAGTCGC TCGTTTGATG AGCTTGCATT GCGAGATTCA CAAGTGGCGG
751 GTAGCGTTTG ATAGGAATTG CTATGCGATG GCAGAAAAGG CGTTTGCAGAA
801 AGCGTTGGGA GCTTTAGAAG AGAGTGTGTA TCGGAGTCTG ACGCAGAGTT
851 ATAGAGATAA ATTTTGGAG AGCGAGAGGG CGAAGATCCC ATGGAATGGG
901 CATATAACCT GGTAAAGAGA TGATCCGAAG AGTGGGTGTG CTGAAAAGAA
951 GCTTCGGGAT GCCGAGGAAC GTTGAAGAAA ATTTAGGAAA GCAGTCTTTT
1001 GGGTAGAAGA AGACGGGGGC TTTGACATCA ATAATCTCCT TGGAGACTGG
1051 GGGACAGTGC TTGATCCTTA TAGACAAGAG AGAATGGACG AGATAACGTT
1101 CCATGAGTTG TATGAAAAAA CTACGTTTTT GAAAAGACTG CACAGAAAGT
1151 GTGCGTTAGC GAAAACAACC TTTGAAAAGA AGAGATCTAA AAAGAATTTG
1201 CAGGCAGTCG AGGAGGCGAA TGCACGTAGG TTGAAATATG TAAGGGATTG
1251 GTATGATCAG GAGTTTCAGA AAGCAGGGGA GAGATTAGAG AAAGTGCATG
1301 CTTTGTATCC TGAGGTTTCA GTCTCTATAA GAGAGAACAA AATACAAGAG
1351 ACGCGCTCTA ATTTAGAGAA AGCCTATGAG GCTATCGAAG AGAAGTATCG
1401 TTGCTGTGTC CGAGAGCAAG AGGACTACTG GAAAGAAGAA GAGAAAAGGG
1451 AAGCGGAGTT TAGGGAGAGG GAAACAAGA TTCCTTCTCC TGAGGAGCTG
1501 GAAAGTTCTT TGGAGCAATT CGACCATGGT TTGAAAAATT TTCTTGAGAA
1551 ATTAATGGAA TTGGAAGGGC ATATCTTAAA ACTTCAGAAA GAAGCCACAG
1601 CAGAGGTGGA GAATAAAATA CTTTCAGATG CAGAGAGCCG CCTTGAGATT
1651 GTATTTGAAG ATGTCAAGGA GATGCCCTGT CGAATTGAGG AGATAGAGAA
1701 GACGCTGCGT ATGGCGGAGC TGCCCCTACT TCCTACGAAG AAGGCGTTTG
1751 AGAAGGCCCTG CTCACAATAT AATAGCTGCG CAGAGATGTT GGAGAAGGTG
1801 AAGCCTTACT GCAAGGAGAG CCTCGCCTAT GTGACTAGCA AAGAGCGTTT
1851 AGTGAGCTTG GATGAAGATT TACGACGAGC CTACACAGAG TGTCAGAAGA
1901 GATTCCAGGG GGATTCGGGT TTGGAGTCGG AAGTAAGAGC CTGTCCGAGG
1951 CAACTGCGAG AGCGGATCCA AGAGTTTGAA ACTCAAGGGC TGGACTTGGT
2001 GGAAAAAGAG TTGCTTTGTG TGAGTAGTAG ATTAAGAAAT ACAGAGTGCG
2051 ATTGTGTATC TGGTGTFAAG AAAGAAGCAC CTCCTGGTAA GAAGTTTAT
2101 GCCCAGTATT ATGATGAGAT TTATCGAGTT AGAGTTCAAT CCCGATGGAT
2151 GACGATGTCT GAGAGATTGA GAGAGGGAGT TCAAGCATGC AACAAGATGT
2201 TGAAGGCAGG CCTAAGCGAA GAAGATAAGG TTCTTAAAGA AGAAGAGTAT
2251 TGGTTGTATC GAGAGGAGAG AAAGAATAAA GAGAAACGTT TGGTTGGTAC
2301 TAAGATAGTA GCAACGCAGC AGCGAGTTGC AGCATTGAA TCCATAGAAG
2351 TTCTTGAGAT TCCTGAGGCC CCAGAGGAGA AACCGAGTTT GCTGGATAAA
2401 GCGCGTTCTT TATTTACTCG CGAGGACCAT ACCTAG
  
```

The PSORT algorithm predicts inner membrane (0.461).

50 The proteins were expressed in *E.coli* and purified as his-tag products (Figure 174: 7200=lanes 2-3; 7236=lanes 4-5; 7268=lanes 6-8; 7375=lanes 9-10; 7388=lanes 11-12). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 174, 175, 176, 177 & 178) and for FACS analysis.

55 These experiments show that cp7200, cp7235, cp7268, cp7375 & cp7388 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 179

The following *C.pneumoniae* protein (PID 4376723) was expressed <SEQ ID 357; cp6723>:

1 MATSVAPSPV PESSPLSHAT EVLNLPNAYI TQPHPIPAAP WETFRSKLST
 51 KHTLCFALTL LLTLGGTISA GYAGYTGNIW ICGIGLGIIV LTLILALLLA
 101 IPLKNKQTGT KLIDEISQDI SSIGSGFVQR YGLMFSTIKS VHLPELTTQN
 151 QEKTRILNEI EAKKESIQNL ELKITECQNK LAQKQPKRKS SQKSFMRISIK
 201 HLSKNPVILF DC*

The cp6723 nucleotide sequence <SEQ ID 358> is:

1 ATGGCAACTT CCGTAGCCCC ATCACCAGTC CCCGAGAGCA GCCCTCTCTC
 51 TCATGCTACA GAAGTTCTCA ATCTTCCTAA TGCTTATATT ACGCAGCCTC
 101 ATCCGATTCC AGCGGCTCCT TGGGAGACCT TTCGCTCCAA ACTTTCCACA
 151 AAGCATACGC TCTGTTTTGC CTTAACACTA CTGTTAACCT TAGGGGGAAC
 201 GATCTCAGCA GGTACGCAG GATATACTGG AAAGTGGATC ATCTGTGGCA
 251 TCGGCTTGGG AATTATCGTA CTCACACTGA TTCTTGCTCT TCTTCTAGCA
 301 ATCCCCTCTTA AAAATAAGCA GACAGGAACA AAAGTGGATG ATGAGATATC
 351 TCAAGACATT TCCTCTATAG GATCAGGATT TGTTTCAGAGA TACGGGTTGA
 401 TGTTCTCTAC AATTAAAAGC GTGCATCTTC CAGAGCTGAC AACACAAAAT
 451 CAAGAAAAAA CAAGAATTTT AAATGAAATT GAAGCGAAAA AGGAATCGAT
 501 CCAAAATCTT GAGCTTAAAA TTAGTGAGTG CCAAAACAAG TTAGCACAGA
 551 AACAGCCGAA ACGGAATACA TCTCAGAAAT CATTATGCG TAGTATTAAG
 601 CACCTCTCCA AGAACCTGT AATTGTGTTT GATTGCTGA

20 The PSORT algorithm predicts inner membrane (0.6095).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 179A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 179B) and for FACS analysis.

25 These experiments show that cp6723 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 180

The following *C.pneumoniae* protein (PID 4376749) was expressed <SEQ ID 359; cp6749>:

1 MSYYFSLWYL KVQQHFQAAF DFTRSLCSRI SNFALGVIAL LPIIGQLYVG
 51 LDWLLSRIKK PEPSPDQDI VRVEHVVGHD HRSRVEDILK RQRLSLEPRD
 101 EGKVHGDLPD APFF*

The cp6749 nucleotide sequence <SEQ ID 360> is:

1 ATGAGTTATT ACTTTTCTCT TTGGTATCTG AAGGTGCAAC AGCACTTTCA
 51 AGCAGCATTG GATTTTACTC GCTCCCTGTG TTCACGAATT TCTAATTTTG
 101 CTTTGGGAGT GATTGCATTG CTTCTTATTA TTGGGCAGTT GTATGTAGGG
 151 CTGGACTGGC TCCTCTCTAG GATAAAAAAG CCAGAATTTT CTTCCGATGT
 201 GGATCAGATC GTGCGAGTAG AACACGTCGT GGGTCACGAC CATAGAAGTC
 251 GAGTTGAAGA TATTCTAAAG AGACAAAGGC TCTCATTAGA GCCTAGAGAC
 301 GAGGGGAAGG TTCACGGAGA TCTGCCTTCA GCTCCTTTT TTTGA

The PSORT algorithm predicts inner membrane (0.2996).

40 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 180A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 180B) and for FACS analysis.

These experiments show that cp6749 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 181 ,
 Example 182 ,
 Example 183 ,
 Example 184 and
 5 Example 185

The following *C.pneumoniae* protein (PID 4376301) was expressed <SEQ ID 361; cp6301>:

```

1 LNQDLQNVYQ ECQKATGLES EVSAYRDHLR EQITEFETQG LDVIKEELLF
51 VSSTLKSKLS YDPLIADIPC MKFYEEYDYG IDKARVQSRW LEKSERYRKA
101 KKGFOEMLKE GLFKEDQALK KAEYRLRLREK RMNKEKLLIC NKIEAAQQRV
151 QEFGPSDS*
  
```

The cp6301 nucleotide sequence <SEQ ID 362> is:

```

1 TTGAATCAGG ATTTACAAAA TGTATACCAA GAGTGCCAGA AGGCTACAGG
51 TTTAGAATCG GAAGTGAGTG CATATAGAGA TCATCTTAGA GAGCAGATCA
101 CAGAGTTTGA AACTCAAGGG CTGGACGTGA TAAAAGAAGA ACTTCTTTTT
151 GTGAGTAGTA CTCTCAAAAG TAAATTGAGC TATGATCCAT TAATAGCAGA
201 CATTCCTGT ATGAAGTTTT ATGAGGAGTA TTATGATGGC ATTGATAAAG
251 CGAGAGTTCA ATCCCGATGG CTGGAGAAGT CTGAGAGGTA TAGAAAGGCG
301 AAGAAGGGAT TCCAAGAGAT GCTGAAGGAA GGCCTATTCA AAGAAGATCA
351 GGCTTTGAAA AAAGCAGAGT AATAGATTACT TCGAGAGAAG AGAATGAATA
201 401 AGGAGAAGCT TTTGATTTCG AATAAGATAG AAGCAGCTCA GCAGCGAGTC
451 CAAGAATTTG GACCTCGGA TTCATAA
  
```

The PSORT algorithm predicts cytoplasm (0.4621).

The following *C.pneumoniae* protein (PID 4376558) was also expressed <SEQ ID 363; cp6558>:

```

1 MNIPAPQVPV IDEPVVNNTS SYGLSLKSSL RPITYLILAI LAIATLMSVL
25 51 YFCGIISVGT FVLGMLIPLS VCSVLCVAYL FYQSSSIEKT KVFSTSPSPV
101 FFSDEDLNLL LGREEDSVSA IDELLKNFPA DDFRRPKMLP YSNFLDEQGR
151 PNESREEDSH TSKIL*
  
```

The cp6558 nucleotide sequence <SEQ ID 364> is:

```

1 ATGAACATAC CCGCTCCCCA AGTACCAGTC ATAGATGAGC CTGTAGTGAA
30 51 CAACACAAGT AGCTATGGTC TTTCATTGAA AAGTAGTTTA AGACCGATTA
101 CTTATTTGAT TTTAGCTATC TTAGCTATAG CCACACTGAT GTCTGTTCTC
151 TACTTTTGTG GCATCATTAG TGTGGGACG TTTGTTTTGG GCATGCTGAT
201 CCCTCTATCG GTCTGCTCTG TTCTTTGCGT TGCCTATTTA TTCTATCAGC
35 251 AATCTTCTAT AGAAAAGACT AAGGTCTTTT CTATAACCAG TCCTTCAGTA
301 TTTTCTCTG ATGAGGATCT TAATTTACTC TTAGGTCGAG AAGAAGATTG
351 AGTGTCTGCA ATTGATGAAC TTCTTAAGAA CTCTCCAGCT GATGATTTC
401 GTAGGCCGAA GATGCTTCCT TATTCAAATT TTCTAGATGA GCAGGGAAGG
451 CCTAATGAGA GTAGGGAAGA AGACTCTCAT ACTTCCAAGA TCTTATAA
  
```

The PSORT algorithm predicts inner membrane (0.4630).

40 The following *C.pneumoniae* protein (PID 4376630) was also expressed <SEQ ID 365; cp6630>:

```

1 MSMTIVPHAL FKNHCECHST FPLSSRTIVR IAIASLFCIG ALAALGCLAP
51 PVSIVGVSIL AFIAFVILSL VILALIFGEK KLPPTPRIIP DRFTHVIDEA
101 YGLSISAFVR EQQVTLAEFR QFSTALLCNI SPEEKIKQLP SELRSKVESF
45 151 GISRLAGDLE KNNWPIPEDL LSQTCPLYWL QKFISAGDPQ VCRDLGVPRE
201 CYGYWLGPL GYSTAKATIF CKETHHILQQ LTKEDVLLK NKALQEKWDT
251 DEVKAIVERI YTTYTARGTL KTEAGGLTKE TISKELLLS LHGYSFDQLQ
301 LITQLPRDAW DWLCFVDNST AYNLQLCALV GALSSQNLLD ESSIDFDVNL
351 GLYVIQDLKE AVQAFSASDE PKKELGKFL RHLSSVSKRL ESVLRQGLHR
401 IALEHGNARA RVYDVNFVTG ARIHRKTSIF FRD*
  
```

50 The cp6630 nucleotide sequence <SEQ ID 366> is:

```

1 ATGAGCATGA CGATCGTTCC ACATGCTTTA TTAAAAATC ATTGCGAGTG
51 TCATTCTACC TTTCCTTTGA GTTCAAGGAC TATTGTAAGA ATAGCCATTG
101 CCAGCCTCTT TTGTATAGGT GCATTAGCAG CTTTAGGCTG TTTGGCTCCT
151 CCCGTTTCTT ATATTGTTGG GAGTGTTTTA GCTTTTATTG CCTTTGTCAT
55 201 TCTTTCTTTA GTAATTTTAG CTTTGATTTT TGGAGAGAAG AAGCTTCCAC
  
```

5
10
15
20

```

251 CAACACCAAG AATCATTCCCT GATAGATTTA CTCACGTGAT AGATGAAGCT
301 TATGGCCTTT CAATCTCTGC ATTTGTAAGA GAACAGCAGG TAACATTAGC
351 CGAGTTTAGA CAATTTTCTA CTGCCCTGTT GTGTAACATA TCTCCTGAAG
401 AGAAAATCAA ACAATTGCCT TCTGAATTGC GAAGTAAAGT AGAGAGTTT
451 GGTATTAGCA GGCTCGCAGG TGATTTAGAA AAGAATAATT GGCCAATATT
501 TGAAGATCTT TTAAGCCAAA CCTGCCCCGT ATATTGGCTT CAGAAATTTA
551 TATCAGCAGG AGATCCACAA GTTTGTAGAG ACCTAGGTGT CCCTAGAGAA
601 TGTTATGGGT ACTATTGGCT AGGGCCTTTG GGATACAGTA CAGCTAAGGC
651 TACAATTTT TGTAAGAGA CGCATCATAT TCTTCAACAA TTAACGAAAG
701 AGGACGTTCT TTTATTAAA AACAAAGGCTC TTCAAGAGAA ATGGGATACT
751 GATGAAGTCA AAGCAATTGT AGAGCGTATC TACACTACCT ATACGGCAGC
801 AGGAACCTTA AAGACCGAAG CAGGGGGACT TACAAAAGAG ACAATCAGTA
851 AGGAATTGCT ATTTGTTGAGC TTGCATGGCT ATCTTTTGA TCAGCTACAG
901 CTGATCACTC AACTTCCTAG AGATGCTTGG GATTGGCTGT GTTTTGTAGA
951 TAACAGTACC GCATACAACC TTCAGCTTTG TGCTCTTGA GGAGCTTTGT
1001 CATCCCCAAA TCTTCTTGAC GAATCTTCTA TCGATTTGA TGTAACCTA
1051 GGCCCTGATG TGATTCAAGG TCTAAAAGAA GCTGTTCAAG CATTTCTGCG
1101 TTCTGATGAG CCAAGAAAG AACTAGGTAA ATTCTTGTTA AGGCATTGTA
1151 GTTCAGTTTC TAAGCGATTA GAGAGTGTAT TAAGACAGGG TCTTCACAGA
1201 ATAGCTCTAG AGCATGGAAA TGCCAGAGCT AGGGTTTATG ACGTCAATTT
1251 TGTAACAGGA GCTAGAATTC ATAGGAAGAC GAGTATCTTC TTTAAAGACT
1301 AA

```

The PSORT algorithm predicts inner membrane (0.7092).

The following *C.pneumoniae* protein (PID 4376633) was also expressed <SEQ ID 367; cp6633>:

25
30

```

1 MVNIQPVYRN TQVNYSQATQ FSVCPALSL IIVSVVAVL AIVALVCSQS
51 LLSIELGTAL VLVSLILFAS AMFMIYKMRQ EPKELLIPKK IMELIQEHYP
101 SIVVDFIRDQ EVSIYIEIHHL ISILNKTNVF DKAPVVLQEK LLQFGIEKFK
151 DVHPSKLPNF EEILLQHCPL HWLGRLVYPM VSDVTPGTYG YYWCGPLGLY
201 ENAPSLFERR SLLLLKKISF GEFALEDGL KNTWSSSEL VQIRQNLFR
251 YYADKBEVDE ABLNADYEQF DSSLHLIFSH KLS*

```

The cp6633 nucleotide sequence <SEQ ID 368> is:

35
40
45

```

1 ATGGTTAATA TACAGCTGT GTATAGGAAT ACCCAAGTCA ACTATAGTCA
51 GGCTACCCAA TTTTCGGTGT GCCAGCCAGC GCTTAGCCTG ATTATCGTTT
101 CTGTTGTTGC TGCTGTACTC GCTATGTAG CTTTGGTATG CAGTCAATCT
151 CTTTATATCCA TAGAGTTAGG AACTGCTCTT GTTCTAGTTT CTCTTATTCT
201 TTTTGCTTCT GCTATGTTTA TGATTATAAA GATGAGACAA GAACCTAAGG
251 AGTTGCTGAT CCCTAAGAAA ATCATGGAAC TCATCCAAGA ACATTATCCA
301 AGTATTGTTG TTGATTTTAT TAGAGATCAG GAGGTTTCCA TTTATGAGAT
351 ACATCACTTG ATCTCTATTC TTAATAAGAC GAATGTTTTT GACAAAGCAC
401 CAGTATATTT ACAAGAAAAA CTCTTACAGT TTGGCATTTA GAAGTTCAAA
451 GATGTACATC CAAGTAAGCT CCCTAATTTT GAAGAAATTC TTCTACAGCA
501 TTGCCCATTT CATTTGGTTGG GACGCTCTGT ATATCCCATG GTATCGGATG
551 TCACTCCAGG AACCTATGGA TACTATTGGT GTGGTCTTTT AGGACTGTAC
601 GAGAACGCTC CCTCTCTTTT TGAACGTCGA TCTCTCTAT TGTTAAAGAA
651 AATTAGCTTT GGAGAGTTTG CTCTTTTAGA AGATGGTCTC AAGAAAAACA
701 CGTGGAGTTC TTCGGAATTC GTTCAAATCA GACAAAACCT TTTTACAAGA
751 TATPATGCTG ATAAAGAAGA GGTAGATGAA GCAGAGTTAA ACGCTGATTA
801 CGAACAGTTT GATTCCTTCC TTCACCTTAT TTTTCTCAC AAGCTCTCTT
851 GA

```

50 The PSORT algorithm predicts inner membrane (0.7283).

The following *C.pneumoniae* protein (PID 4376642) was also expressed <SEQ ID 369; cp6642>:

55

```

1 MATISPISLT VDHPLVDTKK KSCSNFDKIQ SRILLITAIF AVLVTIGTLL
51 IGLLLNIPVI YFLTGISFIA VVLSNFILYK RATLLKPRA CGKHKEIRPK
101 RVSTNLQYSS ISIAINRSKE NWEHQPKDLQ NLPAPSALLT DNPYEIWKAK
151 HSLFSLVSLP PGGNPEHLLI SASENLGKTL LIETTSQNAF ISSYVDTTTPS
201 PKSLNLEAIQ ETRVEINTEL PAGDSGERLY WQPDFRGRVF LPQIPTTPEA
251 IYQYYYALYV TYIQTAINTN TQIIQIPLYS LREHLYSREL PPQSRMQQSL
301 AMITAVKYMA ELHPEYPLTI ACVERSLAQL PQESIEDLS*

```

The cp6642 nucleotide sequence <SEQ ID 370> is:

60

```

1 ATGGCTACAA TCTCACCCAT ATCTTTAACT GTAGATCATC CCCTAGTAGA

```



```

51  CACTAAAAA  AAATCCTGCA  GCAACTTTGA  TAAGATTCAG  TCTCGAATTC
101 TATTGATTAC  TGCAATCTTT  GCTGTCTTAG  TTAATAAGAG  GACCCACTTT
151 ATTGGTTTGC  TTTTAAATAT  TCCTGTTATC  TATTTCTCTA  CAGGAATTTT
201 ATTTATGTCT  GTTGTCTCTA  GCAACTTTAT  CCTTTATAAA  CGAGCAACCA
251 CCCTCTTAAA  ACCGCGTGCT  TGTGGCAAAC  ACAAGAAAT  AAAACCAAAA
301 AGGGTCTCCA  CCAACCTACA  GTATTCTTCT  ATCTCTATCG  CAATCAATCG
351 TTCTAAAGAA  AACTGGGAAC  ACCAACCCAA  GGACCTACAG  AATCTCCCG
401 CACCCTCTGC  ATTACTCACA  GATAACCCCT  ACGAGATATG  GAAAGCTAAA
451 CATTCACGTG  TTTCCCTAGT  ATCCCTCCTA  CCGGGAGGCA  ATCCAGAACA
501 TCTCTTAATT  TCAGCTTCCG  AAAATTTAGG  AAAGACTCTG  TTAATGAAG
551 AAACCTCGCA  AAATGCGCCT  ATATCCTCCT  ACGTAGATAC  CACTCCCTCC
601 CCAAATCCT  TGCTCAATGA  GGCAATTCAG  GAAACCAGGG  TAGAATAAAA
651 TACAGAACTC  CCTGCGGGAG  ATTCAGGAGA  ACGTTTATAC  TGGCAACCCG
701 ATTTCCGAGG  CCGCGTCTTC  CTCCACAAA  TACCAACAAC  TCCTGAAGCC
751 ATCTACCAAT  ACTACTATGC  ACTCTATGTC  ACTTATATCC  AGACTGCGAT
801 CAATACGAAC  ACCCAAATTA  TCCAAATCCC  TTTATACAGC  TTGAGGGAGC
851 ATCTCTATTC  TAGAGAATTG  CCCCAGCAAT  CAAGAATGCA  ACAATCTTTG
901 GCTATGATTA  CAGCAGTAAA  ATACATGGCC  GAGCTGCACC  CAGAATATCC
951 GCTAACTATT  GCTTGTGTTG  AAAGATCCTT  AGCCCAACTA  CCTCAAGAAA
1001 GTATTGAGGA  TCTCTCTTAG

```

The PSORT algorithm predicts inner membrane (0.5288).

The proteins were expressed in *E.coli* and purified as GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 181-185) and for FACS analysis.

- 25 These experiments show that cp6301, cp6558, cp6630, cp6633 and cp6642 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from their sequences alone.

Example 186

The following *C.pneumoniae* protein (PID 4376389) was expressed <SEQ ID 371; cp6389>:

```

30 1 MSEVKPLFLK NDSFDLATQR FQNLINMLQE QAEIYNEYEE KNARVQNEIK
51 EQKDFVKRCI EDFEARGLGV LKEELASLTR DFHDKAKAET SMLIEPCIC
101 FYYSIHQEEQ RQRQERLQKM AERYRDCKQV LEAVQVEQKD MISSRVVVDD
151 SYFEEEEKEEQ KVDNRKKEQD *

```

The cp6389 nucleotide sequence <SEQ ID 372> is:

```

35 1 ATGTCAGAAG TGAAGCCTTT GTTTTAAAG AATGACTCTT TTGATTTGGC
51 AACTCAGAGA TTCCAGAATC TAATTAACAT GCTACAAGAG CAAGCCGAGA
101 TATATAACGA GTATGAAGAA AAGAATGCTA GGGTTCAGAA TGAGATTAAG
151 GAGCAAAAGG ACTTTGTGAA AAGATGCATA GAGGACTTTG AAGCCAGAGG
201 ACTGGGGGTG CTAAAAGAAG AGCTTGCATC TTTGACGCGT GATTTCCATG
40 251 ATAAAGCAAA AGCAGAGACT TCTATGCTCA TTGAATGTCC TTGTATTGGT
301 TTTTATTATA GTATTCATCA GGAGGAACAA AGGCAAAGGC AAGAAAGGCT
351 TCAAAAGATG GCTGAGCGCT ATAGGGACTG TAAACAAGTC TTGGAGGCTG
401 TCCAGGTGGA GCAAAAAGAT ATGATATCTT CTAGAGTCGT TGTCGATGAC
45 451 AGCTACTTTG AAGAAGAAA AGAAGAACA AAGGTGGATA ACAGAAAGAA
501 AGAACAGGAC TAG

```

The PSORT algorithm predicts cytoplasm (0.3193).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 186A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 186B) and for FACS analysis.

These experiments show that cp6389 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 187

The following *C.pneumoniae* protein (PID 4376792) was expressed <SEQ ID 373; cp6792>:

```

5      1 VLQEHFFLSE DVITLAQQLL GHKLITTHEG LITSGYIVET EAYRGPDDKA
      51 CHAYNYRKTO RNRAMYLKGG SAYLYRCYGM HLLNVVTGP EDIPHAVLIR
     101 AILPDQGKEL MIQRRQWRDK PPHLLTNGPG KVCQALGISL ENNRQRLNTP
     151 ALYSISKEKIS GTLTATARIG IDYAQEYRDV PWRFLSPED SGKVL*
```

The cp6792 nucleotide sequence <SEQ ID 374> is:

```

10      1 GTGCTACAAG AACATTTTTT TCTATCGGAA GATGTAATTA CACTAGCGCA
      51 ACAGCTTTTA GGACATAAAC TCATCACAAC ACATGAGGGT CTGATAACTT
     101 CAGGTTACAT TGTAGAAACC GAAGCGTATC GTGGCCCTGA TGACAAAGCA
     151 TGCCACGCCT ACAACTACAG AAAAATCAG AGGAACAGAG CGATGTACCT
     201 GAAAGGAGGC TCTGCTTACC TCTACCGTIG CTATGGCATG CATCACCTAT
     15 251 TGAATGTTGT CACTGGACCT GAGGACATTC CCCATGCCGT CCTGATCCGG
      301 GCCATCCTTC CTGATCAAGG CAAAGAACTT ATGATCCAAC GCCGCCAATG
     351 GAGAGATAAA CCCCCACACC TTCTCACCAA TGGACCCGGA AAAGTGTGCC
     401 AAGCTCTAGG AATCTCTTTG GAAAACAATA GGCAACGCCT AAATACCCCA
     451 GCTCTCTATA TCAGCAAAGA AAAAATCTCT GGGACTCTAA CAGCAACTGC
     20 501 CCGGATCGGC ATCGATTATG CTCAAGAGTA TCGTGATGTC CCATGGAGAT
     551 TTCTCCTATC CCCAGAAGAT TCGGGAAAAG TTTTATCTTA A
```

The PSORT algorithm predicts cytoplasm (0.180).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 187A; lanes 2-4). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 187B) and for FACS analysis.

These experiments show that cp6792 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 188

The following *C.pneumoniae* protein (PID 4376868) was expressed <SEQ ID 375; cp6868>:

```

30      1 MVETVLHNFQ RYLSKYLYRV FRFPCKRKTFF LSSHRVLARP SFPVDYCPGK
      51 IYDLQEIYEE LNAQLFQGA LRLQIGWFRK ATRKGSVVL GLFHENEQLI
     101 RIHRSLDRQE IPRFFMEYLV YHEMVHSVVP REYSLSGRSI FHGKKFKEYE
     151 QRFPLYDRAV AWEKANAYLL RGYKKRVGGG YGRA*
```

The cp6868 nucleotide sequence <SEQ ID 376> is:

```

35      1 ATGGTTGAAA CAGTACTTCA TAATTTCCAA CGTTATCTGA GCAAGTATCT
      51 CTATAGGGTA TTTTCGCTTCC CATGTCGTAA AAAGACGTTT CTATCTTCGC
     101 ACAGGGTTCT TGCTCGTCCT TCATTCACAG TAGACTACTG TCCGGGAAAG
     151 ATCTATGATT TGCAGGAGAT CTATGAGGAA TTGAATGCGC AGTTATTTC
     201 AGGTGCACTG CGTTTACAGA TTGGTTGGTT CGGAAGGAAA GCTACCAGAA
     40 251 AAGGCAAGAG TGTGTCTTGT GGATTGTTTC ATGAAAATGA ACAGTTAATT
      301 CGAATTCATC GTTCTTTAGA TCGGCAGGAA ATCCAAGAT TTTTATGGA
     351 ATATCTTGTG TATCATGAAA TGGTTCATAG TGTAATCCCT AGAGAGTATT
     401 CTCTATCGGG GCGTTCGATT TTTTATGTTA AAAAGTTTAA AGAATACGAA
     451 CAACGTTTCC CTTTGTATGA TCGTGTCTGT GCTTGGGAAA AGGCAACGC
     45 501 TTATTTATTG CGAGGTATA AAAAAGAGT AGGTGGAGGA TATGCGAGG
     551 CATAG
```

The PSORT algorithm predicts bacterial cytoplasm (0.325).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 188A; lanes 2-3). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 188B) and for FACS analysis.

These experiments show that cp6868 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 189

The following *C.pneumoniae* protein (PID 4376894) was expressed <SEQ ID 377; cp6894>:

```

1  MYKRCVLDKI LKGI VAGSLI LLYWSSDLLE RDIKSIKGNV RDIQEDIREI
51  SRVVKQQQTS QAI PAAPGVM LAPKLVRDEA FALLFGDPSY PNLLSLDPYK
101 QQTLPELLGT NFHPHGILRT AHVGKPENLS PFNGFDYVVG FYDLICIPSLA
151 SPHVGYKEEF SPDLAVKIEE HLEDGSGDK EFHIYLRPNV FWRPIDPKAL
201 PKHVQLDEVF QRPHPVTAHD IKFFYDAVMN PYVATMRAVA LRSCYEDVVS
251 VSVENDLKL VVRWKAHTVIN EEGKEERKVL YSAFSNTLSL QPLPRFVYQY
301 FANGEKIIED ENIDTYRTNS IWAQNFTMHW ANNYIVSCGA YYFAGMDEK
351 IVFSRNPDFY DPLAALIDKR FVYFKESTDS LFQDFKTGKI DISYLPNNQR
401 DNFYSFMKSS AYNKQVARGG AVRETVSADR AYTYIGWNCF SLFFQSRQVR
451 CAMNMAIDRE RIIEQCCLDGO GYTISGPFAS SSPSYNKQIE GWHYSPEEAA
501 RLLEEEGWID TDGDGIREKV IDGVIVPFRF RLCYVKSVT AHTIADYVAT
551 ACKEIGIECS LLGLDMADLS QAFDEKNFDA LLMGWCLGIP PEDPRALWHS
601 EGAMEKGSAN VVGPHNEEAD KIIDRLSYEY DLKERNRLYH RFHEIHEEA
651 PYAPLFSRHC SLLYKDYVKN IFVPTHRTDL IPEAQDET VN VTMVWLEKKE
701 DPCLSTS*

```

The cp6894 nucleotide sequence <SEQ ID 378> is:

```

1  ATGTATAAAA GATGTGTGCT AGATAAAATT TTAAAGGGGA TTGTGCGCCGG
25  TTCTTTAATT TTGTTATACT GGTCCCTCAGA CCTACTTGAA AGAGACATTA
101 AGTCGATAAA AGGTAACGTA AGAGATATTC AAGAAGACAT TCGTGAAATC
151 TCACGCGTAG TGAACAACA GCAGACATCA CAAGCTATCC CTGCGGCACC
201 TGGGGTGATG CTCGCTCCTA AGCTCGTCAG AGACGAAGCT TTTGCTCTAC
251 TCTTTGGAGA TCCTAGTTAT CCTAATTTAC TTTCCCTAGA CCCCTATAAA
301 CAGCAGACTC TTCTGAACT TCTAGGAACA AATTTCCACC CTCATGTGTAT
351 CCTACGCACT GCCCATGTGC GAAAACCCGA AAATCTGAGC CCTTTTAAATG
401 GCTTTGATTA TGTCGTGGGC TTTTACGATC TCTGTATTC TAGTTTAGCT
451 TCTCCCCACG TAGGGAAATA CGAAGAAATT TCTCCAGATC TCGCTGTGAA
501 AATAGAAGAA CATCTTGTG AAGATGGTTC TGGGGATAAA GAGTTTCACA
551 TCTATCTGAG GCCGAATGTT TTTGGCGTC CTATAGATCC TAAGGCCCTT
601 CCAAAACACG TTCAGTTAGA CGAAGTATTT CAACGTCCTC ATCCTGTGAC
651 AGCTCATGAT ATTAAGTTT TCTACGACGC TGTATGAAC CCTTATGTAG
701 CAACCATGCG AGCAGTGGCT CTGCGCTCTT GTTATGAAGA TGTGGTTTCT
751 GTCTCAGTAG AAAACGATTT AAAATTAGTA GTCAGATGGA AAGCACACAC
801 GGTAATCAAT GAAGAAGGAA AGGAAGAGCG CAAAGTGCTC TACTCTGCAT
851 TTTCTAATAC CTTAAGCTTG CAGCCCTCC CTAGATTGT ATATCAGTAT
901 TTTGCTAACG GGGAAAAAAT CATTGAAGAT GAGAATATCG ATACCTACCG
951 AACCAATTCC ATTTGGGCGC AAAACTTCAC TATGCATTGG GCAAACAACCT
1001 ATATTGTAAG TTGTGGAGCC TACTACTTTG CAGGGATGGA TGATGAGAAA
1051 ATCGTGTTTT CTAGAAATCC TGACTTCTAT GATCCTCTTG CGGCTCTTAT
1101 TGACAAGCGT TTCGTCTATT TTAAGGAAAG CACAGACTCC CTATTTCAAG
1151 ATTTTAAGAC AGGGAAAATA GACATCTCTT ACCTTCCACC CAACCAAGA
1201 GATAATTTCT ATAGTTTAT GAAAAGCTCC GCTTATAACA AACAGGTAGC
1251 TAAGGGAGGA GCCGTCGGTG AAACAGTCTC AGCAGATCGA GCATATACGT
1301 ACATAGGATG GAATTGCTTT TCATTATTTT TCCAAAGCCG ACAGGTGCGC
1351 TGTGCTATGA ACATGGCAAT CGATAGAGAG AGGATTATCG AACAGTGCTT
1401 GGATGGCCAA GGCTATACGA TTAGTGGGCC TTTTGCTTCG AGTTCTCTTT
1451 CTTATAATAA ACAGATCGAA GGGTGGCATT ATTCTCCAGA AGAAGCAGCT
1501 CGTCTCCTGG AAGAAGAGGG ATGGAATAGAT ACCGATGGCG ATGGAATCCG
1551 AGAAAAAGTT ATCGATGGTG TGATTGTCCC GTTCCGTTTC CGTTTATGCT
1601 ATTATGTAAG GAGTGTCACC GCTCATACCA TTGCAGATTA CGTAGCTACT
1651 GCTTGTAAGG AAATCGGAAT CGAGTGTAGC CTTCTAGGAC TAGATATGGC
1701 CGATCTTTTC CAAGCTTTTG ATGAAAAGAA TTTTCGATGCT CTTTAAATGG
1751 GATGGTGTTC AGGAATTCCT CCTGAGGATC CTAGGCTTT ATGGCATCTC

```

5 1801 GAAGGGGCTA TGGAAAAGGG TTCAGCGAAT GTTGTAGGTT TCCATAATGA
 1851 AGAAGCTGAT AAAATCATAG ACAGACTCAG CTACGAATAC GATCTGAAAG
 1901 AACGTAATCG CCTGTACCAC CGTTTCCATG AAATTATTCA TGAGGAAGCT
 1951 CCTTATGCTT TCTTGTCTC ACGACATTGT TCCTTACTTT ATAAGGATTA
 2001 TGTA AAAAAT ATTTTCGTAC CTACACATAG AACAGATTTA ATTCCTGAAG
 2051 CTCAGGATGA GACTGTCAAC GTAAC'TATGG TATGGCTTGA GAAGAAGGAG
 2101 GATCCGTGCT TAAGTACATC CTAA

The PSORT algorithm predicts inner membrane (0.162).

10 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 189A) and also in GST/his form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 189B) and for FACS analysis.

These experiments show that cp6894 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 190

15 The following *C.pneumoniae* protein (PID 4377193) was identified in the 2D-PAGE experiment <SEQ ID 379; cp7193>:

20 1 MKRVIYKTIF CGLTLLTSL SCSLDPKGYN LETKNSRDLN QESVILKENR
 51 ETPSLVKRLS RRSRLRFARR DQTQKDTLQV QANFKTYAEK ISEQDERDLS
 101 FVVSSAAEKS SISLALSQGE IKDALYRIRE VHPLALIEAL AENPALIEGM
 151 KKMQRDRIW NLFLTQLSEV FSQAWSQGI SEEDIAAFAS TLGLDSGTVA
 201 SIVQGERWPE LVDIVIT*

A predicted leader peptide is underlined.

The cp7193 nucleotide sequence <SEQ ID 380> is:

25 1 ATGAAAAGAG TCATTTATAA AACCATATTT TGCGGGTAA CTTTACTTAC
 51 AAGTTTGAGT AGTTGTTCCC TGGATCCTAA AGGATATAAC CTAGAGACAA
 101 AAAACTCGAG GGA CTTAAAT CAAGAGTC TG TACTGAA GAAAACCGT
 151 GAAACACCTT CTCTGTGTTA GAGACTCTCT CGTCGTTCTC GAAGACTCTT
 201 CGCTCGACGT GATCAAAC TC AGAAGGATAC GCTGCAAGTG CAAGCTAACT
 251 TTAAGACCTA CGCAGAAAAG ATTTTCAGAGC AGGACGAAAG AGACCTTTCT
 301 TTCGTTGTCT CGTCTGCTGC AGAAAAGTCT TCAATTTCGT TAGCTTTGTC
 351 TCAGGGTGAA ATTAAGGATG CTTTGTAACG TATCCGAGAA GTCCACCCTC
 401 TAGCTTTAAT AGAAGCTCTT GCTGAAAACC CTGCCTTGAT AGAAGGGATG
 451 AAAAAGATGC AAGGCCGTGA TTGGATTGG AATCTTTTCT TAACACAATT
 501 AAGTGAAGTA TTTTCTCAAG CTTGGTCTCA AGGGGTTATC TCTGAAGAAG
 35 551 ATATCGCCGC ATTTGCCCTCC ACCTTAGGTT TGGACTCCGG GACCGTTGCG
 601 TCCATTGTCC AAGGGGAAAG GTGGCCCGAG CTTGTGGATA TAGTGATAAC
 651 TTAA

The PSORT algorithm predicts periplasmic (0.925).

This shows that cp7193 is an immunoaccessible protein in the EB and that it is a useful immunogen.

40 These properties are not evident from the protein's sequence alone.

It will be appreciated that the invention has been described by way of example only and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE II – sequences of the primers used to amplify Cpn genes.

Orf ID	N-terminus final primer	C-terminus final primer
CP0014P	GCGTC CCG GGT CATATG AAGTCTTCTTCCCA	GCGT CTC GAG ATGAAAGAGTTTTGCG
CP0015P	GCGTCCCGGGTCATATG TCAGTCTGTTTTCTGA	GCGT CTC GAG GAATTGGTATTTTGCTC
CP0016P	GCGTCCCGGGTCATATG GCCGATCTCACATTAG	GCGT CTC GAG GTCCAAGTTAAGGTAGCA
CP0017P	GCGT CCG GGT CATATG GGTATCAAGGGAAGT	GCGT CTC GAG AAATCCGAATCTTCC
CP0019P	GCGTCCCGGGTCAT ATGCAAGACTCTCAAGACTATAG	GCGT CTC GAG AAATCGGTATTTTACCC
CP6260P	GCGTC CCG GGT GCTAGCACTACGATTCTTTAAGCC	GCGT CTC GAG AAAACGAAATTTGCTTC
CP6397P	GCGTC CCG GGT CATATG TCACTCTCTGTAAATAACA	GCGT CTC GAG ATGAAAGAGAGTCTCTCG
CP6456P	GCGTC CCG GGT CATATG TCACTCTCTGTAAATAACA	GCGT CTC GAG CTGACCATCTCTCTGTT
CP6466P	GCGTC CCG GGT CAT ATG TGCAAGGAGTCCAGT	GCGT CTC GAG ATTTTCTCTAGCATAACG
CP6467P	GCGTC CCG GGT CAT ATG TGTTCCTCATCCCAA	GCGT CTC GAG TAGTTTTCTATAAAACGAAAGTCT
CP6468P	GCGTC CCG GGT CAT ATG TGCTCTCTCTACTCTTC	GCGT CTC GAG GGGGAAATAGGTATATTGTA
CP6469P	GCGTC CCG GGT CAT ATG AGCTGCTCAAAGCAA	GCGT CTC GAG ACTTAAGATATCGATATTTTGA
CP6552P	GCGTC CCG GGT CAT ATG TGCCATAAGGAAGATG	GCGT CTC GAG ACCATTGTCTTGAGTCAT
CP6567P	GCGTC CCG GGT CAT ATG ACCTCACCGATCCCC	GCGT CTC GAG AGAAGCCGATAGAGGC
CP6576P	GCGTC CCG GGT CAT ATG ACTGAAAGGTTAAAGAAAG	GCGT CTC GAG GAA CATGCCCTCAA
CP6727P	GCGTC CCG GGT CATATGCTACATCCACTAATGGC	GCGT CTC GAG GAAAGAAACGAGTTCC
CP6729P	GCGTC CCG GGT CAT ATGGCAGATGCTTCTTTATC	GCGT CTC GAG GAATGAGTATCTTAGCC
CP6731P	GCGTC CCG GGT CATATGGCTGTGTGAAATCAAT	GCGTC CAT GGC GGC CGC GAACCTGGAACCTACCTCC
CP6736P	GCGTC CCG GGT GCT AGCGTAGAAGTTATCATGCCTT	GCGTC CAT GGC GGC CGC AAATCGTAATTTGCTTC
CP6737P	GCGT GGA TCC CAT ATG GAGACTAGACTCGGAGG	GCGT CTC GAG AAATGTGATTTTAGTCC
CP6751P	GCGTC CCG GGT GCT AGC AATGAAGGTCTCCAAT	GCGT CTC GAG AAATCTCATTTACTCTGC
CP6752P	GCGTGA ATT CAT ATGTTCCGGATGACTCCT	GCGT CTC GAG GAAATTTAAGGTACTTCTCTG
CP6753P	GCGTC CCG GGT GCT AGCACTCCCTACTCTCATAGAG	GCGT CTC GAG AAACCTAAAGGTCGTTT
CP6767P	GCGTC CCG GGT CAT ATG ATAAACAAATAGGCCGT	GCGT CTC GAG TTCGTAAGCAACTTCAGA
CP6829P	GCGTC CCG GGT CAT ATG AAGCAGATGCGTCTTT	GCGTC CAT GGC GGC CGC GAACTAAGGGAGAGGC
CP6830P	GCGTC CCG GGT CAT ATG GATCCCGCTCTGTT	GCGTC CAT GGC GGC CGC GAATACAAACCGATCC
CP6832P	GCGTC CCG GGT CAT ATG CATAAAGTAATAGTTTTCATTT	GCGT CTC GAG TAAACTAGAAAAAGTCGTC
CP6848P	GCGTC CCG GGT CAT ATG TCATCAATCTACATCCC	GCGT CTC GAG AAGCGGAGCTATTTTAC
CP6849P	GCGTC CCG GGT GCT AGC AGCGGGGTATAGAG	GCGT CTC GAG ATACACGTGGGTATTTTC
CP6850P	GCGTC CCG GGT CAT ATG TGCCGATTTAGAT	GCGT CTC GAG CTGTTGCACTTGCC
CP6854P	GCGTC CCG GGT GCT AGC TCAATAGCTATTGCAAG	GCGT CTC GAG TTATCGAAATGTCTTTG
CP6879P	GCGTC CCG GGT CAT ATG GCAACACCCGCTCAA	GCGTC CAT GGC GGC CGC TCCTTGAATTTGCTCTTGC
CP6894P	GCGTC CCG GGT CAT ATG TATAAAGATGTGTGCTAGA	GCGT CTC GAG GGATGTACTTAAGCAGC
CP6900P	GCGTC CCG GGT CAT ATG AAGATAAAATTTTCTTGAAG	GCGT AAG CTT GGGAGAGCAGTACCG
CP6952P	GCGTC CCG GGT CAT ATG CTCTCGGATCAATATATAGG	GCGT CTC GAG TCGAATTTCTTTTATAGC
CP7034P	GCGTC CCG GGT CAT ATG AAAAACAAGGTATATCAATG	GCGT AAG CTT AAACGCTGAAATTTATACC
CP7090P	GCGTC CCG GGT CAT ATG TGTAGCCTTTCCOCT	GCGT CTC GAG GCGTGCATGAATCTTA
CP7091P	GCGTC CCG GGT CAT ATG GAAGAATTAGAAGTTGTGT	GCGT CTC GAG TAGTGTCTCTTTTATCGGT
CP7170P	GCGTC CCG GGT CAT ATG CTAGGGGCTGGAAACC	GCGT AAG CTT AAACCTGCAGACCTGACG
CP7228P	GCGTC CCG GGT CAT ATG ACTGCTGTTCTTATTCTTACA	GCGT CTC GAG ATCTGAAAGCGGAGG
CP7249P	GCGTC CCG GGT CAT ATG ATCCATCCCTTACC	GCGT CTC GAG ATCAGGTGCTGAGACTT
CP7250P	GCGTC CCG GGT CAT ATG AATCTTTCAACAGGTCT	GCGT CTC GAG ATTTTCTCTAGAGAGACTCTC
CP0018P	GTGCGT CATATG GCAACCACTCCAATA	ACTCGCTA GCGGCCGC TAATGAGTCCCCAG
CP6270P	GTGCGT CATATG AATTTATTAGGAGCTGCT	ACTCGCTA GCGGCCGC AAATTTGATTTTGCTACC
CP6735P	GTGCGT CATATG GCAGCACAAGTTGTATAT	ACTCGCTA GCGGCCGC TGCGGTAGAATGTATC
CP6998P	GTGCGT CATATG TTGCCTGTAGGGAAC	ACTCGCTA GCGGCCGC GAATCTGAAGTACCAGA
CP7033P	GTGCGT CATATG GTTAATCCTATTTGGTCCA	ACTCGCTA GCGGCCGC TTGGAGATAACCAAGATATA
CP7287P	GTGCGT CATATG TTACACAGCTCAGAACTAGA	ACTCGCTA GCGGCCGC GAAATAATACGGATACCA
CP0010P	GTGCGT CATATG GCAACTGCTGAAATATA	GCGT CTCGAG GAATTTGGAACCTTACCC
CP0468P	GTGCGT GCTAGC ATTTTATTGACAACTCTAT	GCGT CTCGAG AAATGTGCAATGACTCT
CP6272P	GTGCGT CATATG TTGACTCATCAAGAGCT	GCGT CTCGAG GAAGGGAGGTTTTTTAGGT
CP6273P	GTGCGT CATATG ACATATCTGGAAGCTC	ACTCGCTA GCGGCCGC CTCACAAATTTTATG
CP6362P	GTGCGT CATATG CCCTTTGATATTACTTATTATACA	GCGT CTCGAG TCGTTTCCAAATCCA
CP6372P	GTGCGT CATATG AAACAACACTATTCTCTAAATA	GCGT CTCGAG TTCTTGTGTTTTTTCT
CP6390P	GTGCGT CATATG CGAGAGGTGCCTAAG	ACTCGCTA GCGGCCGC TCTCTAGACAGCCTT
CP6402P	GTGCGT CATATG AATGTTGCGGATCTCTTT	GCGT CTCGAG GAAGGGGTTGGCCGT
CP6446P	GTGCGT CATATG TGTAATCAAAAGCCCTCTT	GCGT CTCGAG GGGCTGAGGAGGAAC
CP6520P	GTGCGT GCTAGC AAACACTACCTATCATTTTCT	GCGT CTCGAG CAGAAAGGCTTTTCTTT
CP6577P	GTGCGT CATATG AATTTAGGCTATGTTAATTTA	GCGT CTCGAG GTTTTGTTTTTTGAAGA
CP6602P	GTGCGT CATATG GCAGCATCAGGAGCA	GCGT CTCGAG TGACCAAGGATAGGGTTAG

CP6607P	GTGCGT CATATG CCTCGTGGTGACACTTT	GCGT CTCGAG CGCTGCTCTTGCTC
CP6615P	GTGCGT CATATG TGCTCTCAAAAACGACAA	GCGT CTCGAG TGAAGAGCGCCATC
CP6624P	GTGCGT CATATG GATGCGAAAATGGGA	GCGT CTCGAG TCTTTGACATTCAAGAGC
CP6672P	GTGCGT CATATG ATTCTTACCATTGTTAATG	GCGT CTCGAG GTCATACAATTTCCTTATATA
CP6679P	GTGCGT CATATG TGCACCTCACTTAGGCT	GCGT CTCGAG CGAGTAGTTAGCACAAAC
CP6717P	GTGCGT GCTAGC AAGACAATCGTAGCTTCA	ACTCGCTA GCGGCCGC GGCTGGCATATAGGT
CP6784P	GTGCGT GCTAGC AAATCAAGATGTTCTATTGATA	GCGT CTCGAG TCCAAAAACACCTCT
CP6802P	GTGCGT CATATG TGCGTAAGTTATATTAATTCCTT	GCGT CTCGAG CAGTCGGGCTTGTG
CP6847P	GTGCGT CATATG TCGGATCTTTACGAG	GCGT CTCGAG TTTTCTACACTGTGTGTAATAAA
CP6884P	GTGCGT CATATG AATCAGCTGCTTTCT	GCGT CTCGAG AGAGAAGGTAATTGTACC
CP6886P	GTGCGT CATATG TGTCTACTTATTATCTATCTCTAC	GCGT CTCGAG TTCAGAAAAATGGCT
CP6890P	GTGCGT CATATG TCCCCACGACGACAA	GCGT CTCGAG TCCTGCAGCATTTAGC
CP6890P	GTGCGT CATATG TGTGACGTACGGTCTA	ACTCGCTA GCGGCCGC TTCACCTTGATTTCCT
CP6898P	GTGCGT CATATG TGCGATGCAAAAC	ACTCGCTA GCGGCCGC GGAAGTATGCTTAGATATT
CP6899P	GTGCGT CATATG TGCTGTGGTTACTCTATT	ACTCGCTA GCGGCCGC AAAAAGGTATAGTATACCT
CP7005P	GTGCGT CATATG AAAACTGTGATATTGAACA	GCGT CTCGAG CTGAGCTTCTATTCTTATTAT
CP7072P	GTGCGT CATATG CCCATTTATGGGAAA	GCGT CTCGAG GTTGAGCAAGGTTTG
CP7101P	GTGCGT CATATG TATTCTGTGTACAGCAA	GCGT CTCGAG GAAAAATCTTTAGGGAG
CP7102P	GTGCGT CATATG GCCGCTAAAGCAAA	GCGT CTCGAG TGAATAATGAAAGGATGCT
CP7105P	GTGCGT GCTAGC AGTCTATATCAAAAATGGTG	GCGT CTCGAG ATCTTTCATTGTGGTTACT
CP7106P	GTGCGT CATATG AAAGATTGGGGACTCT	GCGT CTCGAG GAATCCTAGGCAATACCTA
CP7107P	GTGCGT GCTAGC AGTATAGTCAGAAATCTGCA	GCGT CTCGAG GAAGCTAAGATTATAGCTACTTT
CP7108P	GTGCGT GCTAGC GCGGCCCTTCCA	ACTCGCTA GCGGCCGC TTTATGTATATGGAACAGATAGG
CP7109P	GTGCGT CATATG GGACATTTTATGTATATTG	ACTCGCTA GCGGCCGC ATCATCAAGGTAGATAAAG
CP7110P	GTGCGT CATATG GGTATTGTCTATGTAAATTACA	GCGT CTCGAG TTCTGATTGGACTCCA
CP7127P	GTGCGT CATATG GTGGCTTTAACGATAGC	ACTCGCTA GCGGCCGC GCAGCCATCGTATTC
CP7130P	GTGCGT CATATG TTCAATATGCGAGG	GCGT CTCGAG CTTCCTATTGTGAACCTTG
CP7140P	GTGCGT CATATG ACAGCCGAGAGCAGCT	GCGT CTCGAG AGCACCCTCAATTTCATTG
CP7182P	GTGCGT CATATG GGATATGTTTCTATGTGATC	GCGT CTCGAG GCTACTAAATCGAATCGA
CP6262P	GTGCGT CATATG ATCCCTGGATTAGTTCA	ACTCGCTA GCGGCCGC TTCCTGGGAGCTTGA
CP6269P	GTGCGT CATATG TACCAGGAGAACTAAGAT	ACTCGCTA GCGGCCGC GATTTTCTTCTTCAGCTC
CP6296P	GTGCGT CATATG GAGGAGGTGCTGAGTAT	ACTCGCTA GCGGCCGC ATGTTTCTTTTACTCTTTCT
CP6419P	GTGCGT CATATG GCTCCAGTCCGTGTT	GCGT CTCGAG AAGTGTTCGTTGGAAGT
CP6601P	GTGCGT CATATG AATAAGCTACTCAATTTCTGT	GCGT CTCGAG GAAAACTGAATTCCTCTCT
CP6639P	GTGCGT CATATG TTAATTTCAAGCAATTCA	GCGT CTCGAG AGGAATAAAACCTCATCT
CP6664P	GTGCGT GCTAGC GTTTTATTTTCATGCTCAA	ACTCGCTA GCGGCCGC CTTAGAAAGACTATTTTCTAAGTA
CP6696P	GTGCGT CATATG TCCGTGATAATGGG	GCGT CTCGAG ATTATCTTCGTAAAGAAAT
CP6757P	GTGCGT CATATG GCAGTTGGTGGCGT	ACTCGCTA GCGGCCGC CTGTCCCTCTGAGAGC
CP6790P	GTGCGT GCTAGC AGTGAACACAAAAATCA	ACTCGCTA GCGGCCGC CTGTATCGTGTATCAATA
CP6814P	GTGCGT CATATG CATGACGCACTTCTAAG	GCGT CTCGAG TACAGCTGCGCGA
CP6834P	GTGCGT CATATG GTTATGGGAACCTATATCG	GCGT CTCGAG TACATTGTATTGATTTCAG
CP6878P	GTGCGT CATATG AACGTCCCTGATTCC	GCGT CTCGAG GCTAGCGGCTCTTTTC
CP6892P	GTGCGT CATATG CAGAAGCATCCTTCTCT	ACTCGCTA GCGGCCGC TCCTCTTTAGGAAATGG
CP6909P	GTGCGT CATATG TCCTCTTTAGGAAATGG	GCGT CTCGAG CAGTGCCAAAGTAGGGA
CP7015P	GTGCGT CATATG GCAGTACGATTAAATTGTTG	GCGT CTCGAG TTTATTGTAGTCTATTTATATTTT
CP7035P	GTGCGT GCTAGC AGCAGAAAAGACAATGA	GCGT CTCGAG ATTTTGAGTGTCTTGCA
CP7073P	GTGCGT CATATG ATTACCATAAATCAGCTG	GCGT CTCGAG TATCCATCGACTTATAGC
CP7085P	GTGCGT GCTAGC TGTATTTCCTTACGTA	ACTCGCTA GCGGCCGC GGATTCTGCATCTCTG
CP7092P	GTGCGT CATATG TCTCCTCTTCTTAAAAAA	GCGT CTCGAG GGAATTCATTACTGACCA
CP7093P	GTGCGT CATATG AAATACCGCTTCACG	GCGT CTCGAG ATTCTGTAGGGCTACGT
CP7094P	GTGCGT CATATG GTACACTTCTCTCATAAACCC	GCGT CTCGAG TAAGTTTGTATTGTGCGTAT
CP7132P	GTGCGT CATATG TTGTTATTAGGGACTTTAGGA	GCGT CTCGAG TTTCCCAACCGCA
CP7133P	GTGCGT CATATG GCTGCGAATGCTC	GCGT CTCGAG TAATTTAAFACTCTTTTGAAAGG
CP7177P	GTGCGT CATATG CCTACTCAAGTTAAAAACAGA	GCGT CTCGAG AAGTTTATATTTTCAGCACTT
CP7184P	GTGCGT GCTAGC CATATAGGATTTTGCCA	GCGT CTCGAG GTACTTAGCAAGCGAT
CP7206P	GTGCGT GCTAGC AAGAAGCTATATCACCCCTA	GCGT CTCGAG CACACCGAGGAAC
CP7222P	GTGCGT CATATG GTAGTTTCAGAAAGAAAAGTC	GCGT CTCGAG ACGTATGCGCAACTG
CP7223P	GTGCGT CATATG GAAGTATTAGACCGCTCT	GCGT CTCGAG CGAGAAAAGCTTCC
CP7224P	GTGCGT CATATG ATGAAGAAAATTCGAAA	ACTCGCTA GCGGCCGC TAAGCATTCACAAATGA
CP7225P	GTGCGT CATATG CATATTTTGTCTTGATCGT	GCGT CTCGAG TCTTTTAACTTAACTCTTGTCTCT
CP7303P	GTGCGT CATATG CTGTCTATTGTTTGTATCC	GCGT CTCGAG AAAATATACGGAACCTCGC
CP7304P	GTGCGT GCTAGC GAAGTTTATAGTTTTCCTC	GCGT CTCGAG TTTTGTATTCCTTAAGAG
CP7305P	GTGCGT CATATG GAAGTTTATAGTTTTCACCCCT	GCGT CTCGAG ACTCCTTGAGAAAGGAA
CP7307P	GTGCGT CATATG CTTAATCATGCTAAAAAGC	ACTCGCTA GCGGCCGC CTCTTTTATTTTAGGAGCT

CP7342P	GTGCGT CATATG AAAAAAAATTATTTTCTACT	ACTCGCTA GCGGCCGC CACACTCTGTTCTCTG
CP7347P	GTGCGT CATATG TTTTCTAAGGATTGACTAA	GCGT CTCGAG CGAAGCAGAAGTCGT
CP7353P	GTGCGT CATATG AATATGCCTGTTCTCTCT	GCGT CTCGAG GGGGCGTAGGTTGTA
CP7193P	GTGCGT CATATG TGTCCCTGGATCCT	ACTCGCTA GCGGCCGC AGTTATCACTATATCCACAAG
CP7248P	GTGCGT GCTAGC CTGGAACATTCTAAACAAGAT	GCGT CTCGAG ACGTAGTTTAAGAGCAGACT
CP7261P	GTGCGT CATATG TGTCTATCTGCCCTACATAG	GCGT CTCGAG TTTTGATGCTTCTTTCA
CP7280P	GTGCGT CATATG GACCAGAAAATTGAAAA	GCGT CTCGAG AGAGGCTCTTCTGAGTGC
CP7302P	GTGCGT CATATG AATTTCATTGTAGTGTAGT	GCGT CTCGAG GAACAGTTCGATTGTG
CP7306P	GTGCGT CATATG CTTCTTTATCAGGGCA	ACTCGCTA GCGGCCGC TTCTTCAGGTTTCAGG
CP7367P	GTGCGT GCTAGC CGTTATGCCGAGGTC	GCGT CTCGAG TTCGTGCATTGTGGT
CP7408P	GTGCGT CATATG TTGAAAATCCAGAAAA	GCGT CTCGAG ATTCATTTTCGGAAGAG
CP7409P	GTGCGT CATATG AGACGTTATCTTTTCATGGT	GCGT CTCGAG CCCTTTGCTCTTTACATAG
CP6733P	GTGCGT ACTAGT TGTCACCTACAGTCACTAG	GCGT CTCGAG GAATCGGAGTTTGGA
CP6728P	GTGCGT ACTAGT AAGTCCTCTGTCTCTTG	GCGT CTCGAG GAAACAAACTTAGAGCCC

TABLE III - Proteins with best results in FACS analysis

cp number	Molecular Weight (kDa)		Fusion type
	Theoretical	Western Blot	
6260	97.5	94; 70	GST
6270	87.5	-	GST
6272	78.0	90	GST
6273	58.6	74; 64; 50	GST
6296	31.1	-	GST
6390	88.9	102	GST
6456	42.5	89; 67; 45	GST
6466	57.5	59; 56	His
6467	59.0	67	GST
6552	28.4	50; 27	GST
6576	86.0	79; 70; 62; 45	GST
6577	17.3	12	GST
6602	43.4	53; 42; 34	GST
6664	54.5	104; 45	GST
6696	47.9	95; 53	GST
6727	130.0-142.9	123; 61; 39	His
6729	94.8	multiple bands	GST
6731	95.5	97	GST
6733	97.1	104	His
6736	100.1	98; 93; 66; 60	GST
6737	101.2	multiple bands	GST
6751	100.2	95; 71	GST
6752	102.1	97; 48	His
6767	29.1	28	GST
6784	32.9	35	GST
6790	71.3	multiple bands	His
6802	29.7	-	GST
6814	29.6	28	GST

6830	177.4	174; 91; 13	GST
6849	57.3	multiple bands	GST
6850	7.4-9.4	61; 14; 8	GST
6854	42.2	-	GST
6878	40.4	-	GST
6900	28.0	-	GST
6960	25.6	75; 35	GST
6968	34.6	83; 53; 35	GST
6998	39.3	multiple bands	GST
7033	68.2	multiple bands	GST
7101	113	105	GST
7102	63.4	-	GST
7105	29.2	30	GST
7106	39.5	72; 46	GST
7107	71.4	67; 31	His
7108	35.9	35	GST
7111	46.1	51	GST
7132	17.9	57; 47; 17	His
7140	36.2-29.8	50; 38; 34	GST
7170	34.4	77; 33	GST
7224	39.4	40	GST
7287	167.3	180	GST
7306	50.1	50	GST

TABLE IV – FACS-positive proteins not found in *C.trachomatis*

cp7105	cp6390
cp7106	cp6784
cp7107	cp6296
cp7108	

TABLE V – Proteins identified by MALDI-TOF following 2D electrophoresis

cp6270	cp6733	cp6900
cp6552	cp6736	cp6960
cp6576	cp6737	cp6998
cp6577	cp6752	cp7033
cp6602	cp6767	cp7108
cp6664	cp6784	cp7111
cp6727	cp6790	cp7170
cp6728	cp6830	cp7287
cp6729	cp6849	cp7306

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
2. A protein having 50% or greater sequence identity to a protein according to claim 1.
3. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
4. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.
5. A nucleic acid molecule according to claim 4, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318,

320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.

- 5 6. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.
- 10 7. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 4 to 6.
- 15 8. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 4 to 7.
9. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 4 to 8 under high stringency conditions.
- 20 10. A composition comprising a protein or a nucleic acid molecule according to any preceding claim.
11. A composition according to claim 10 being a vaccine composition.
12. A composition according to claim 10 or claim 11 for use as a pharmaceutical.
- 25 13. The use of a composition according to claim 10 in the manufacture of a medicament for the treatment or prevention of infection due to *Chlamydia* bacteria, particularly *Chlamydia pneumoniae*.

1/169

FIGURE 1

FIG. 1A

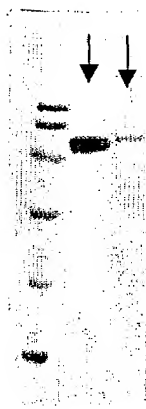


FIG. 1B

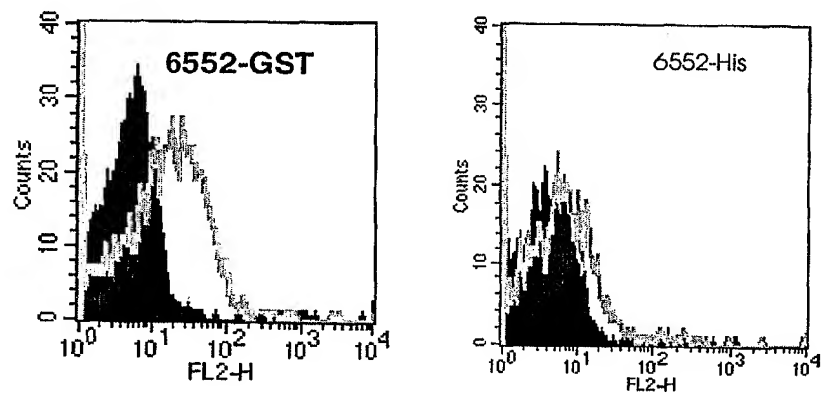
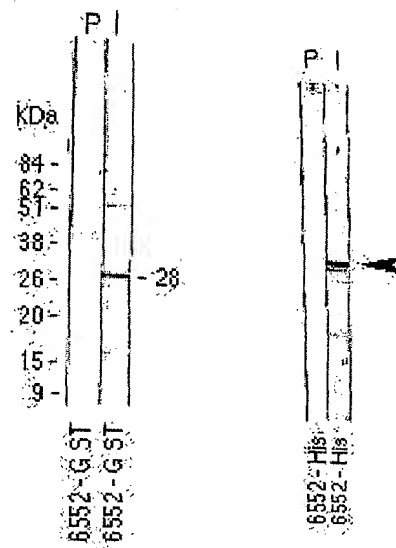
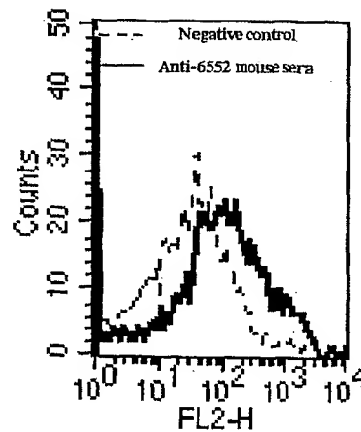


FIG. 1C



2/169

FIGURE 2

FIG. 2A

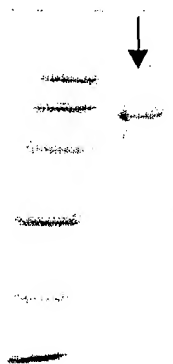


FIG. 2B

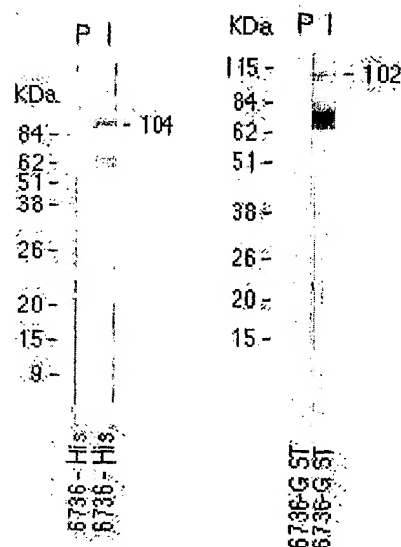


FIG. 2C

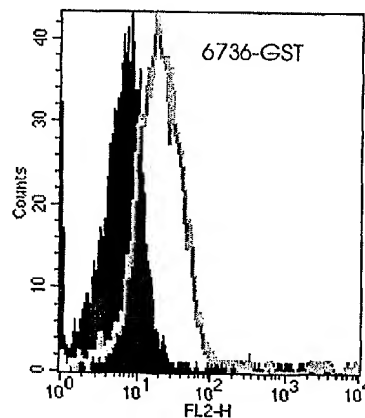
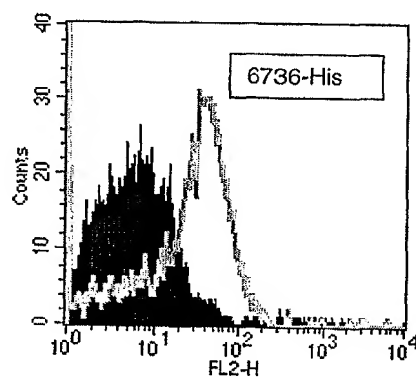


FIGURE 3

FIG. 3A

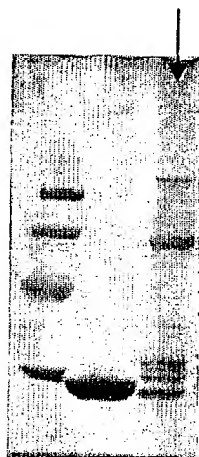


FIG. 3B

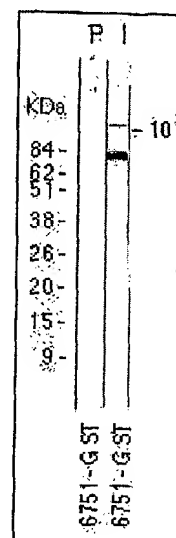
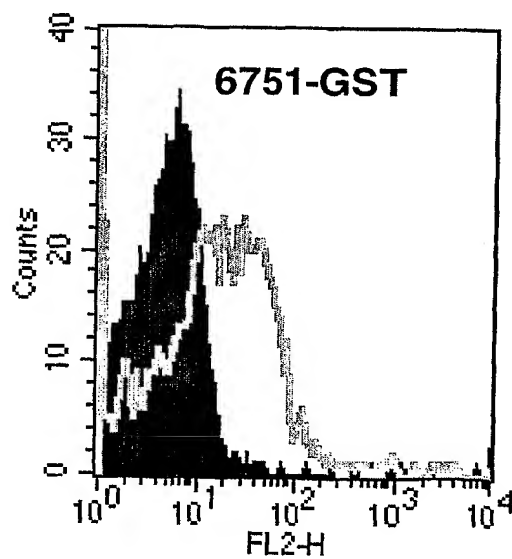


FIG. 3C



4/169

FIGURE 4

FIG. 4A



Fig. 4B

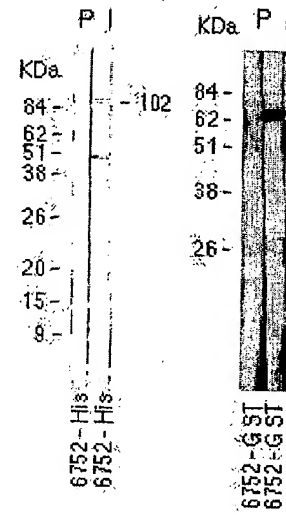


FIG. 4C

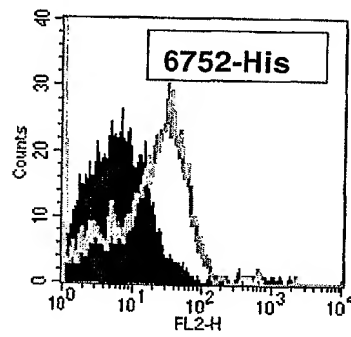


FIGURE 5

Fig. 5A



Fig. 5B

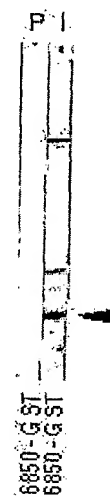
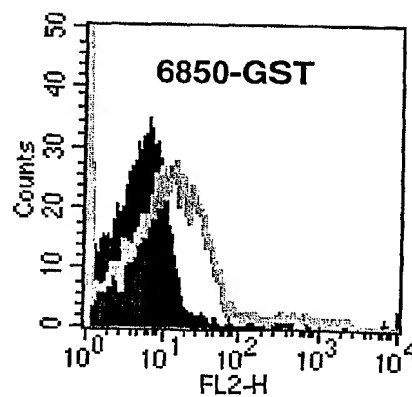


Fig. 5C



6/169

FIGURE 6

FIG. 6A

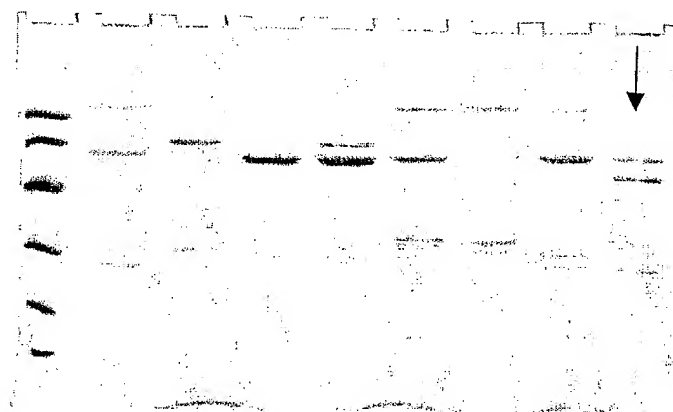


FIG. 6B

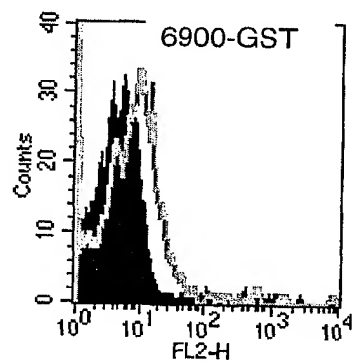
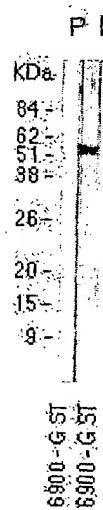


FIG. 6C



7/169

FIGURE 7

FIG. 7A

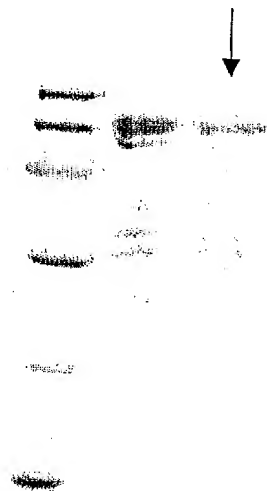


FIG. 7B

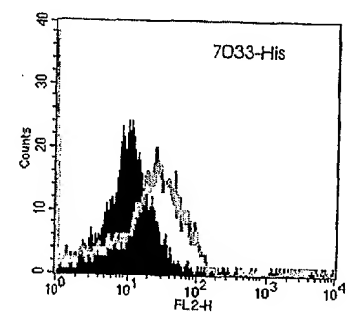
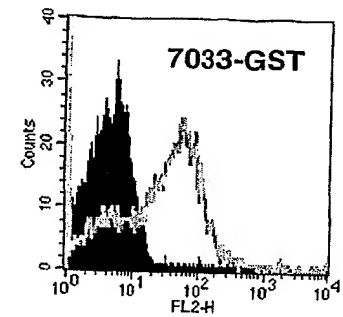
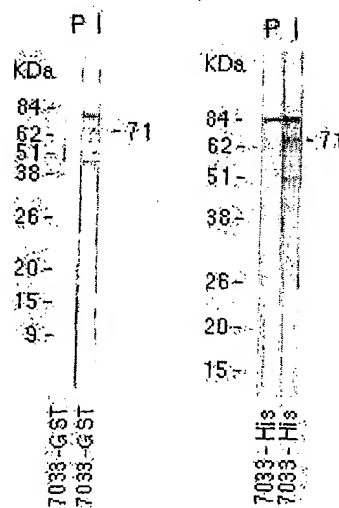


FIG. 7c



8/169

FIGURE 8

Fig. 8A

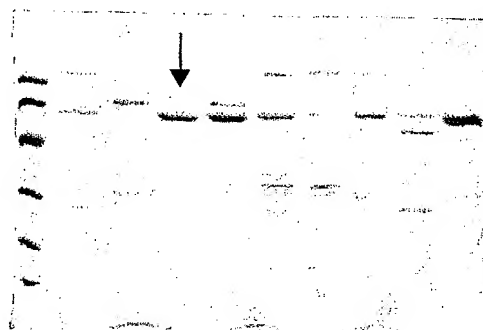


Fig. 8B

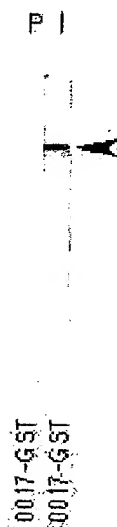
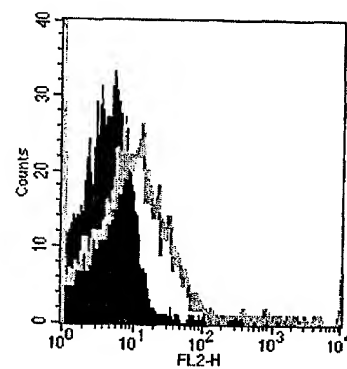


Fig. 8C



9/169

FIGURE 9

Fig. 9A

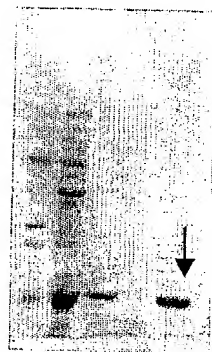


Fig. 9B

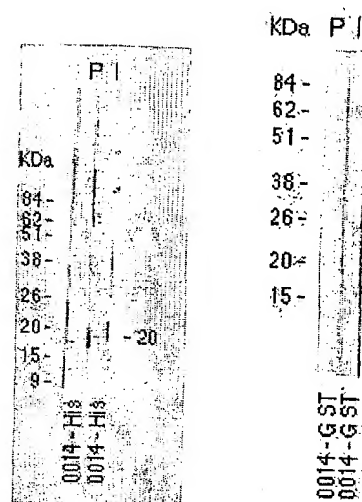
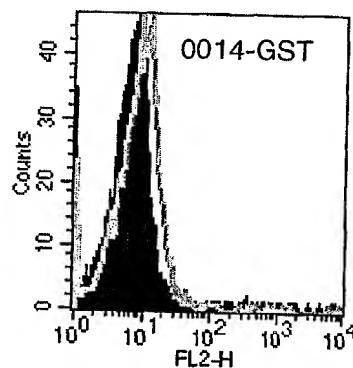


Fig. 9C



10/169

FIGURE 10

FIG. 10A

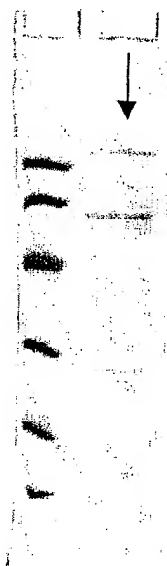
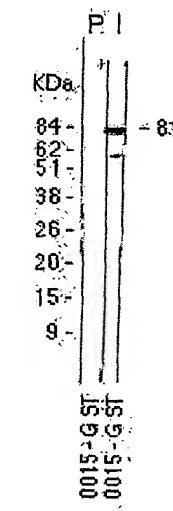


FIG. 10B



11/169

FIGURE 11

Fig. 11A

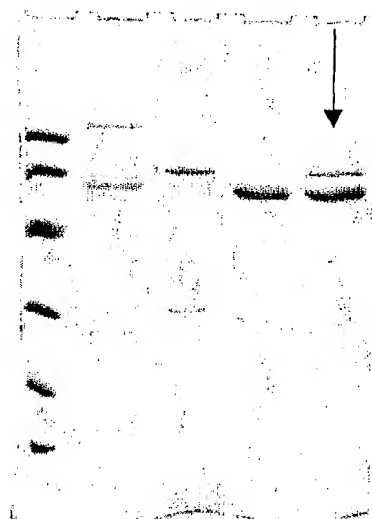


Fig. 11B

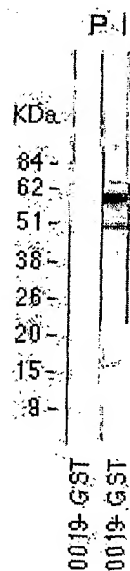
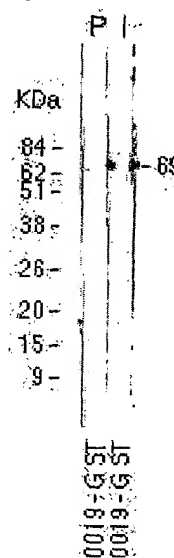


Fig. 11C



12/169

FIGURE 12

Fig. 12A

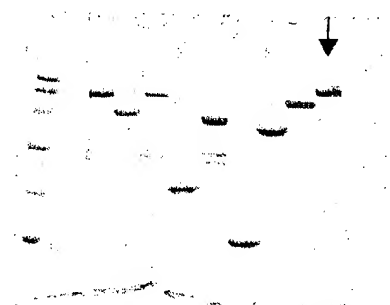


Fig. 12B

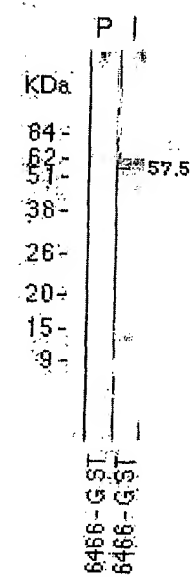
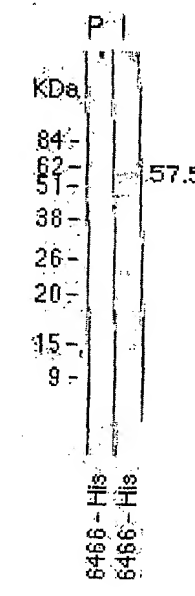


Fig. 12C



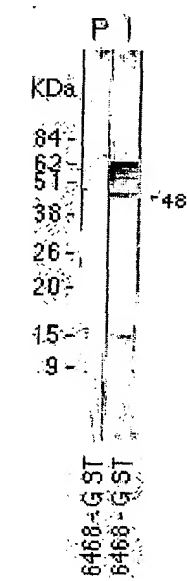
13/169

FIGURE 13

FIG. 13A



FIG. 13B



14/169

FIGURE 14

FIG. 14A

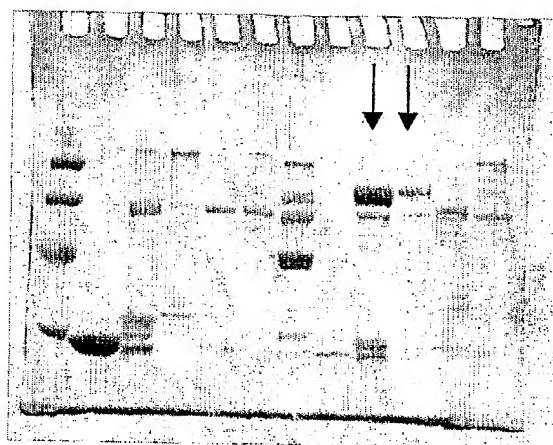
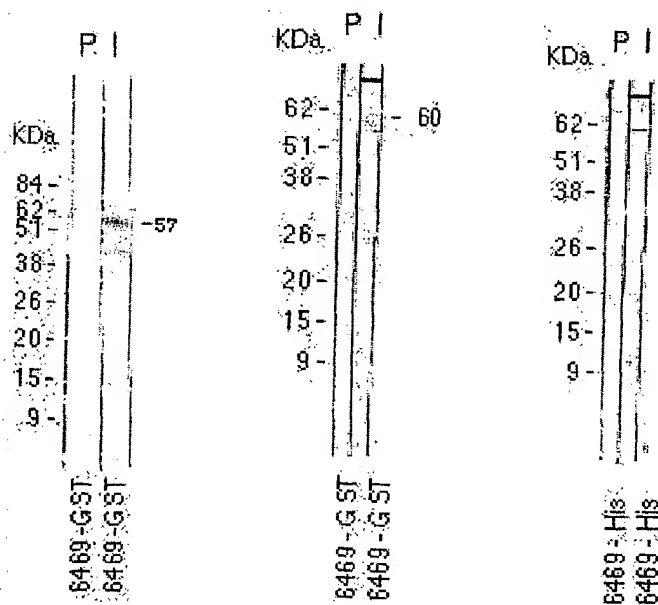


FIG. 14B



15/169

FIGURE 15

Fig. 15A



Fig. 15B

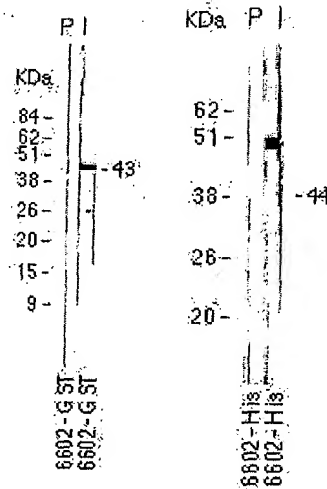
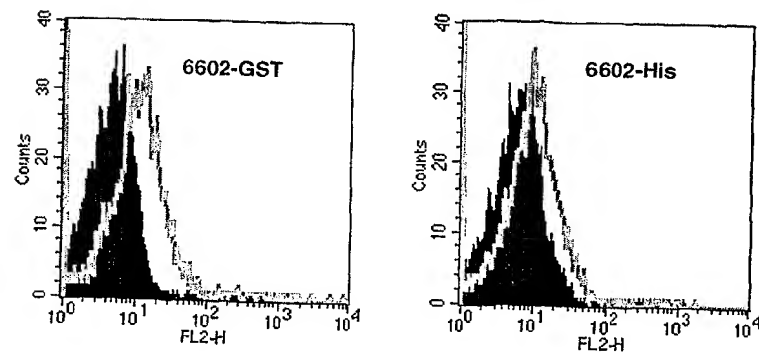


Fig. 15C



16/169

FIGURE 16

FIG. 16A



Fig. 16B

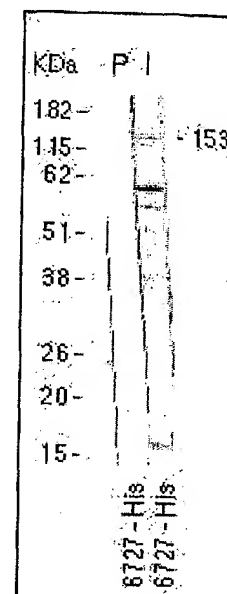
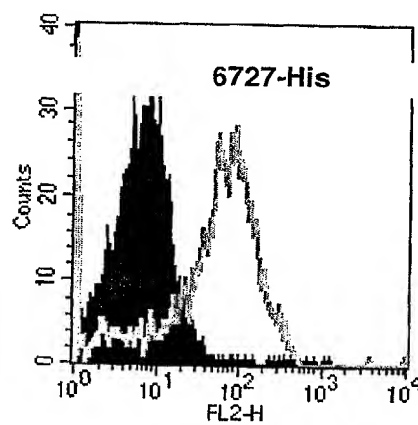


FIG. 16C



17/169

FIGURE 17

FIG. 17A



FIG. 17B

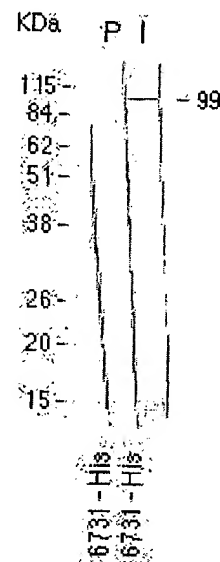


FIG. 17C

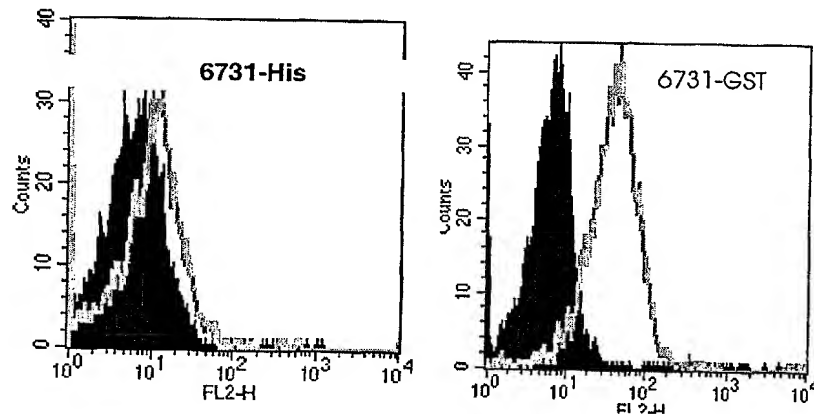


FIGURE 18

FIG. 18A

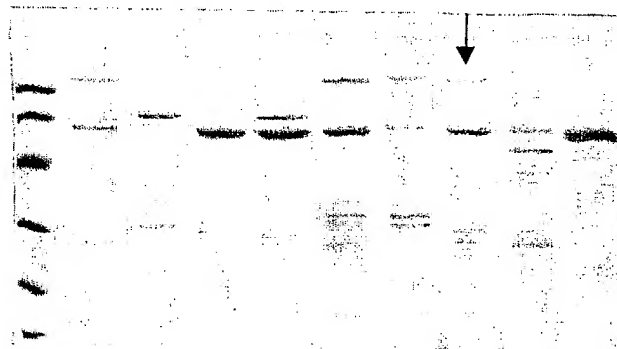


FIG. 18B

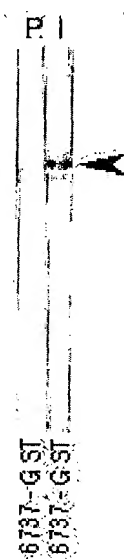


FIG. 18C

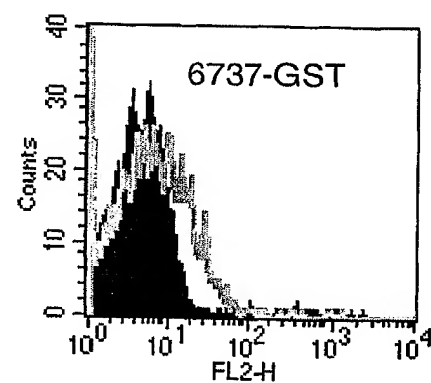


FIGURE 19

FIG. 19A

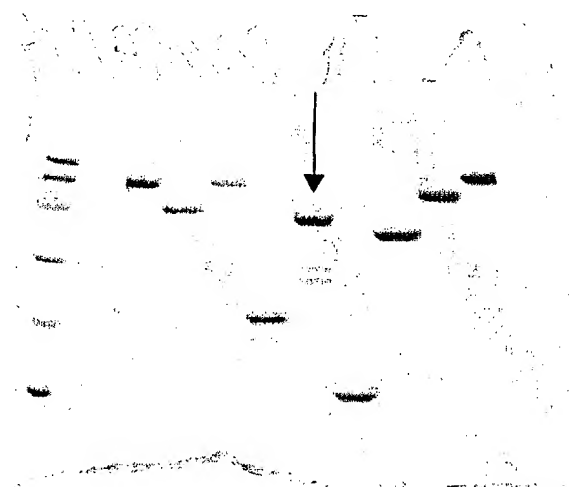
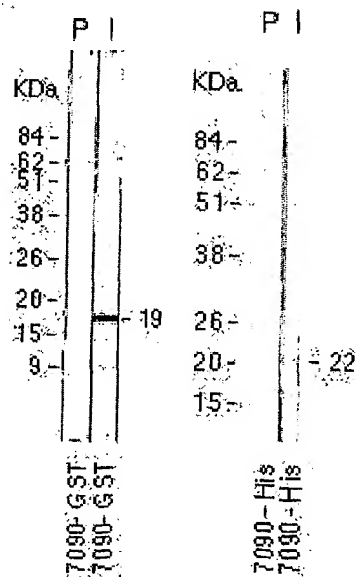


FIG. 19B



20/169

FIGURE 20

Fig. 20A

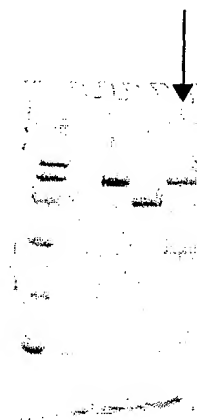
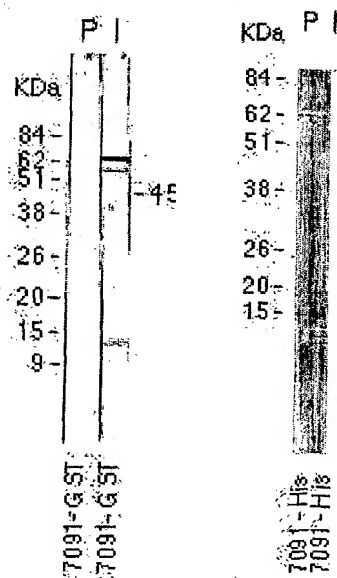


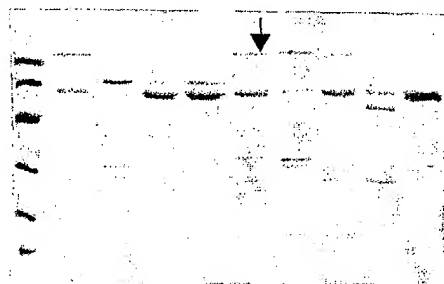
Fig. 20B



21/169

FIGURE 21

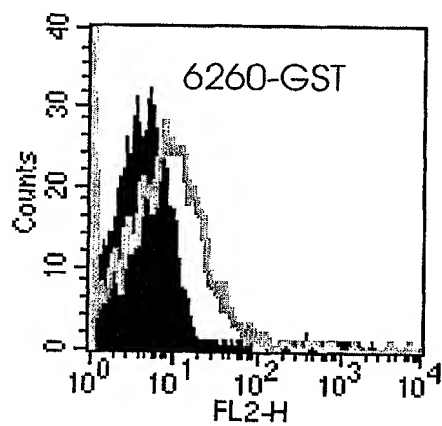
**FIG.
21A**



**FIG.
21B**



**FIG.
21C**



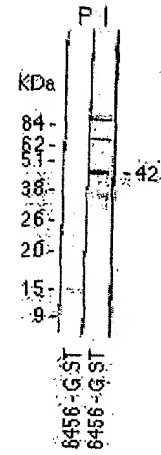
22/169

FIGURE 22

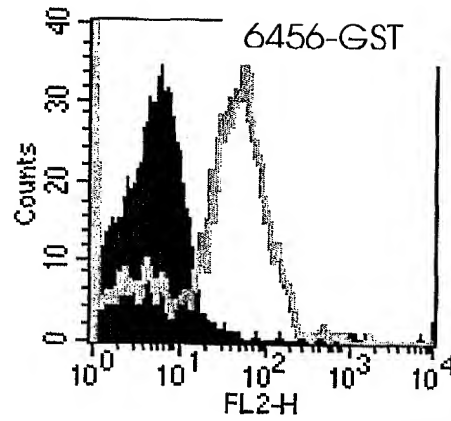
**FIG.
22A**



**FIG.
22B**



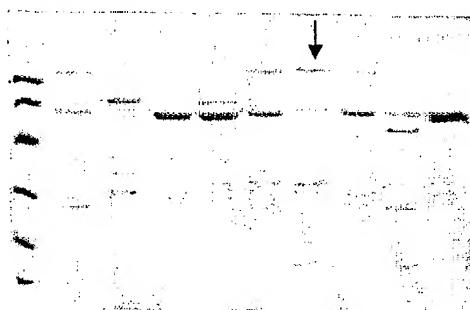
**FIG.
22C**



23/169

FIGURE 23

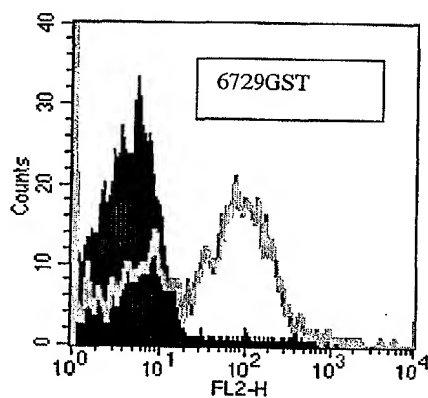
**FIG.
23A**



**FIG.
23B**



**FIG.
23C**



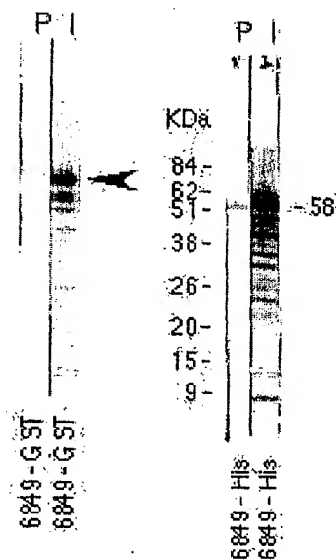
24/169

FIGURE 24

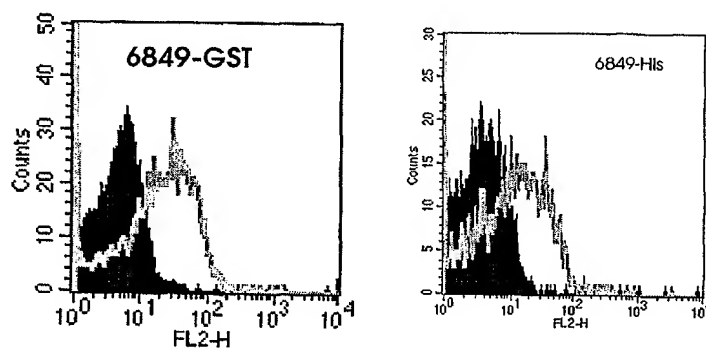
**FIG.
24A**



**FIG.
24B**



**FIG.
24C**



25/169

FIGURE 25

Fig. 25A

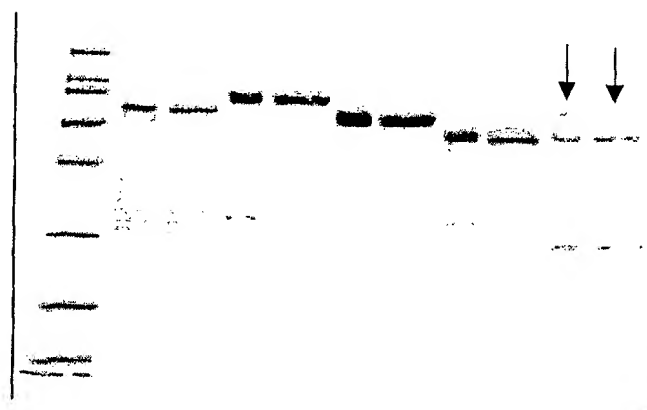


Fig. 25C

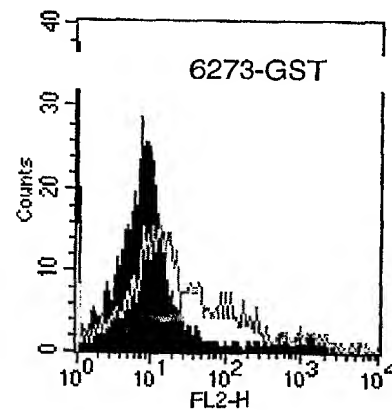
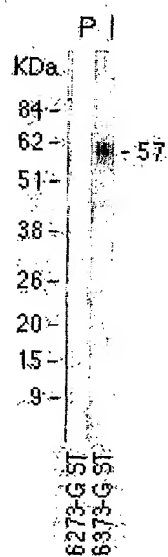


Fig. 25B



26/169

FIGURE 26

Fig. 26A

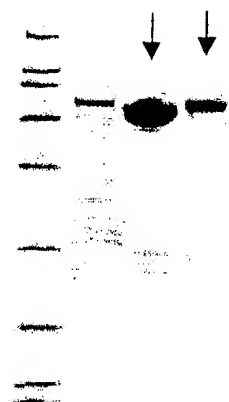
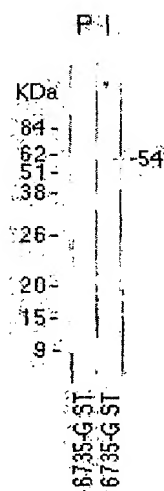


Fig. 26B



27/169

FIGURE 27

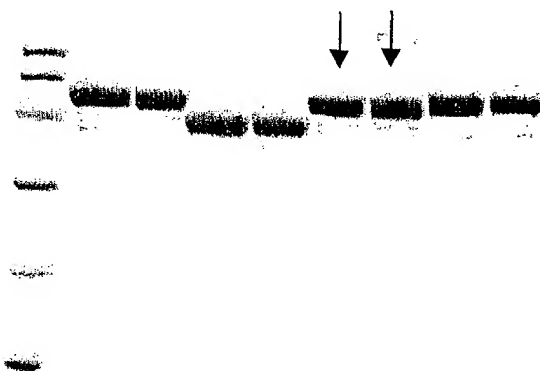


Fig. 27A

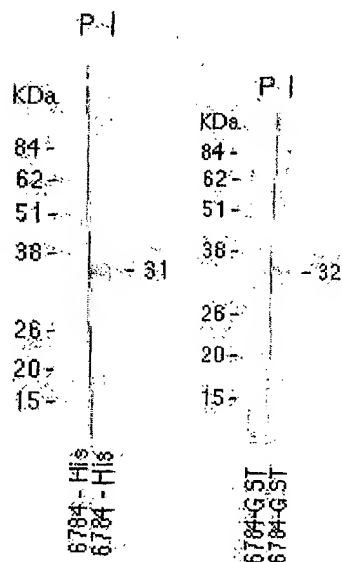
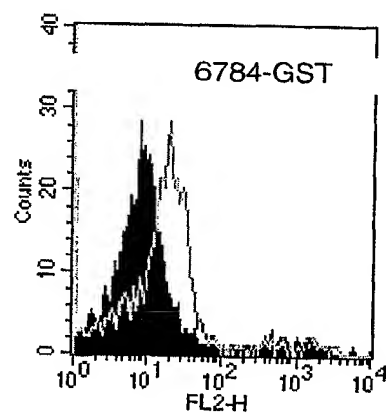


Fig. 27B

Fig. 27C



28/169

FIGURE 28

FIG. 28A

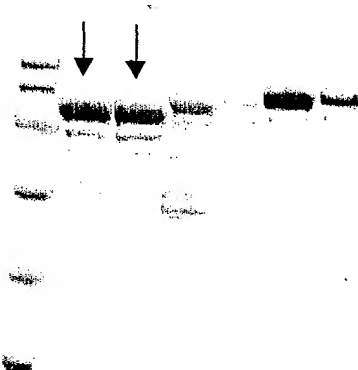


FIG. 28B

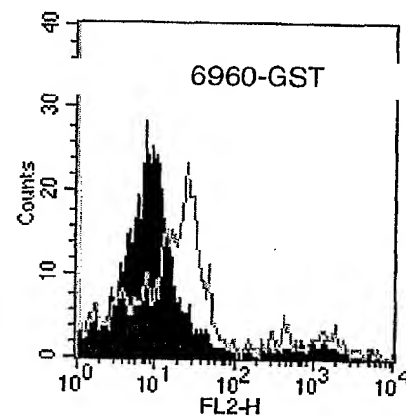
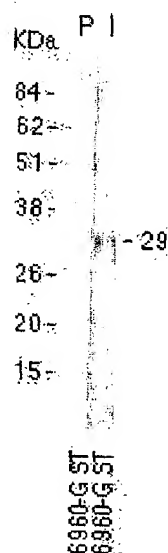
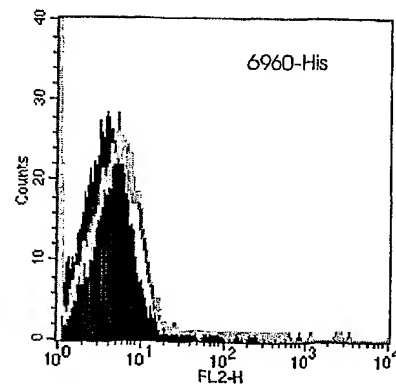


FIG. 28C



29/169

FIGURE 29

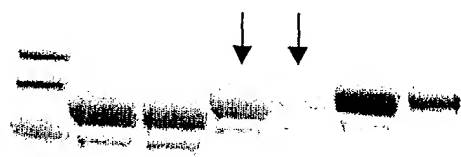


Fig. 29A

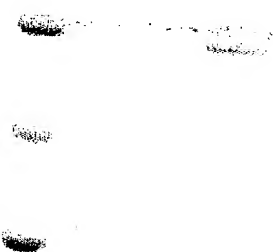


Fig. 29B

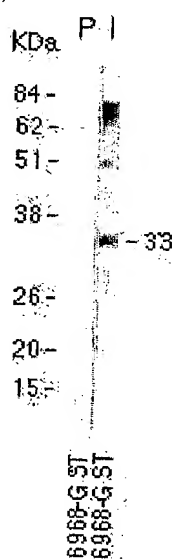
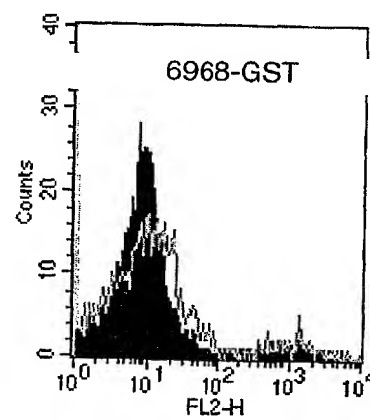


Fig. 29C



30/169

FIGURE 30

Fig. 30A

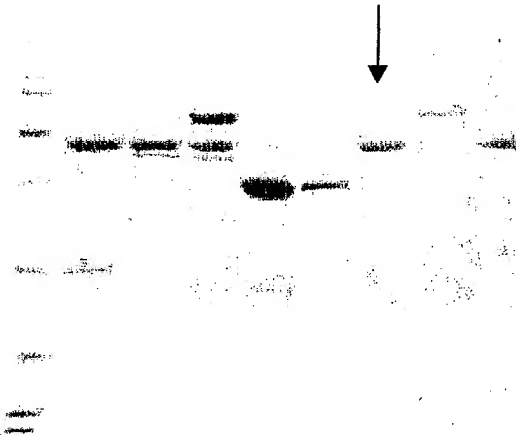


Fig. 30B

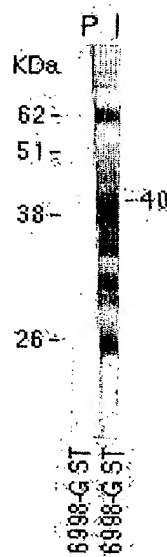
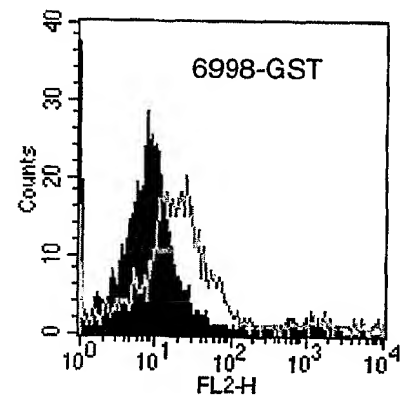


Fig. 30C



31/169

FIGURE 31

Fig. 31A

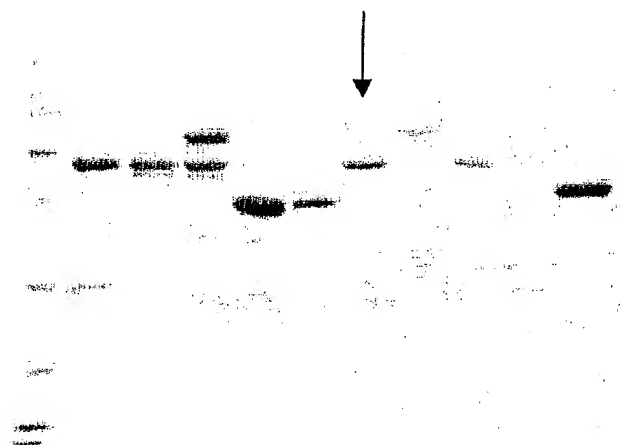
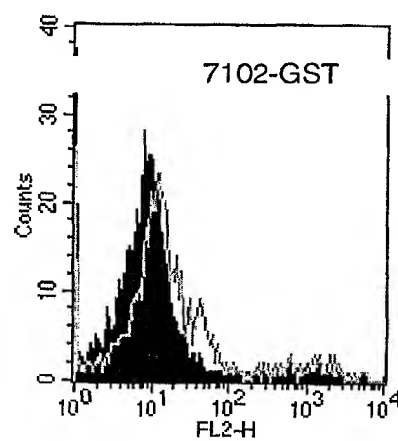


Fig. 31B



32/169

FIGURE 32

Fig. 32A

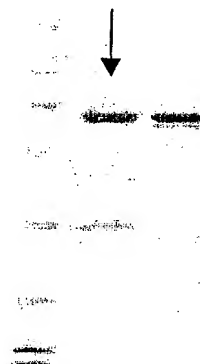


Fig. 32B

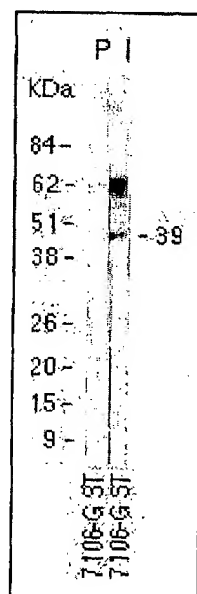
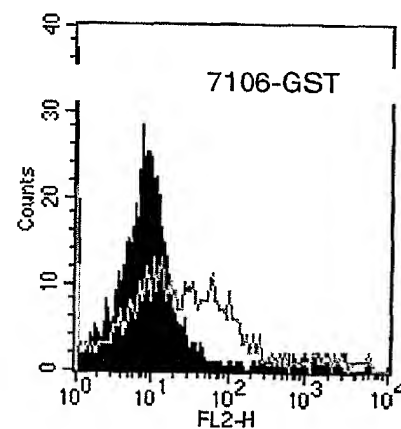


Fig. 32C



33/169

FIGURE 33

Fig. 33A

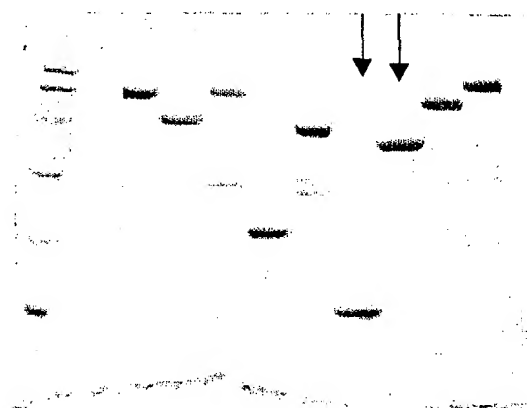
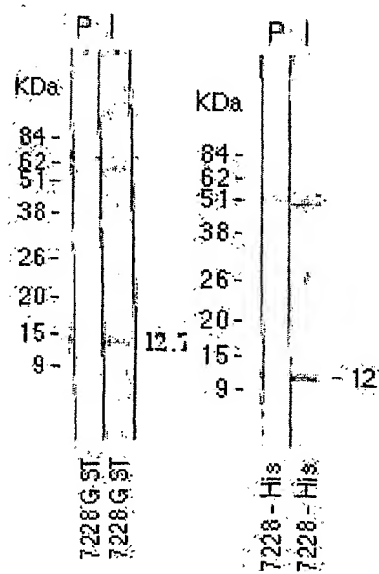


Fig. 33B



34/169

FIGURE 34

FIG. 34A

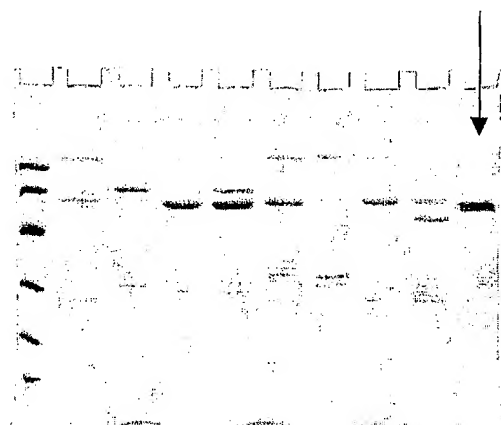


FIG. 34B

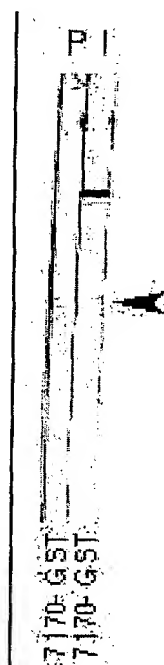
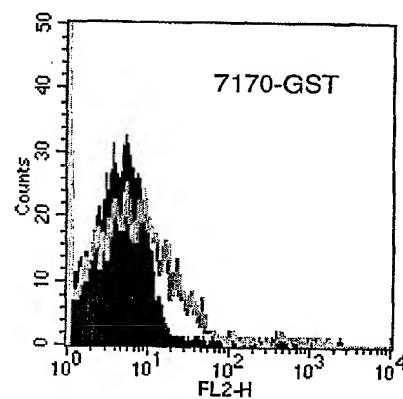


FIG. 34C



35/169

FIGURE 35

Fig. 35A

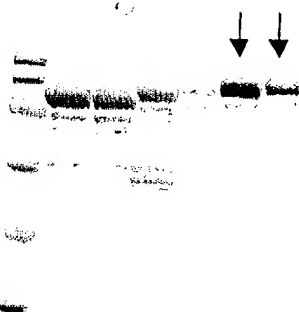


Fig. 35B

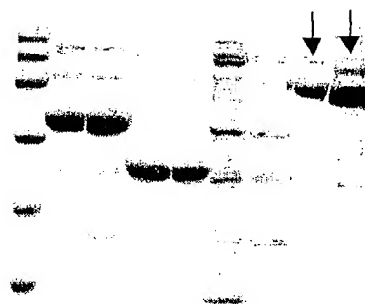
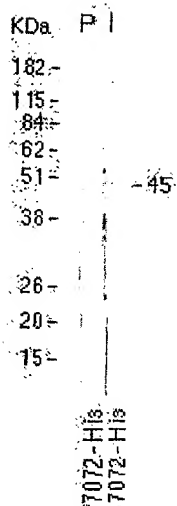


Fig. 35C



36/169

FIGURE 36

Fig. 36A

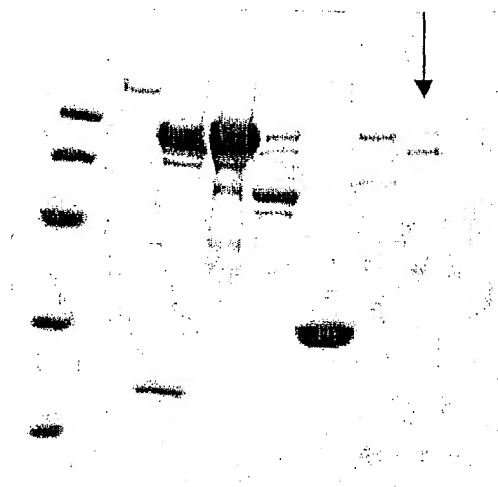
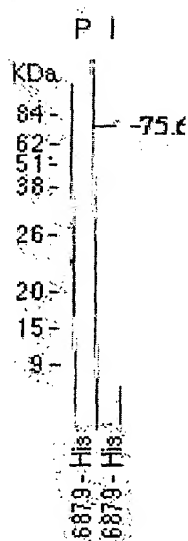


Fig. 36B



37/169

FIGURE 37

FIG. 37A

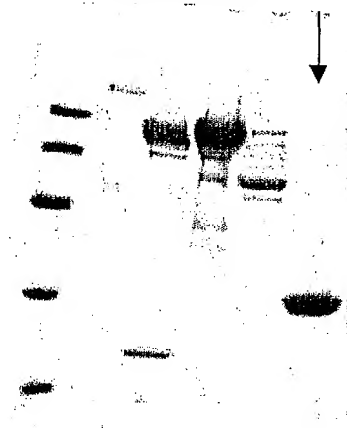


FIG. 37C

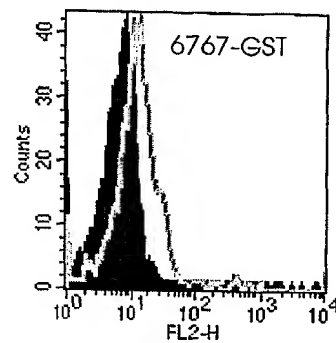


Fig. 37B

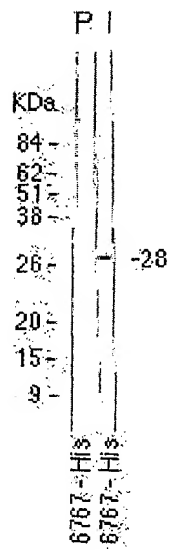
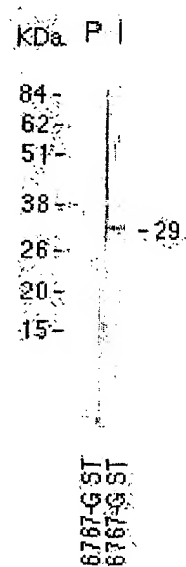


Fig. 37D



38/169

FIGURE 38

FIG. 38A

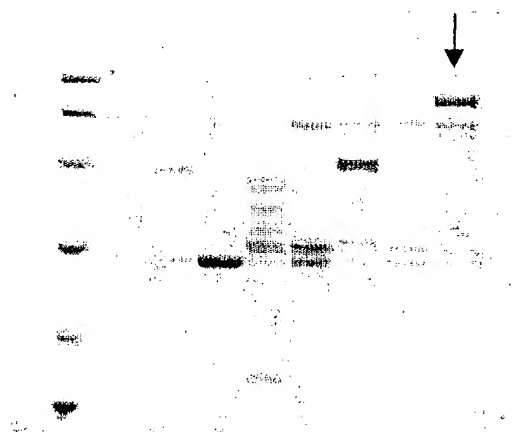
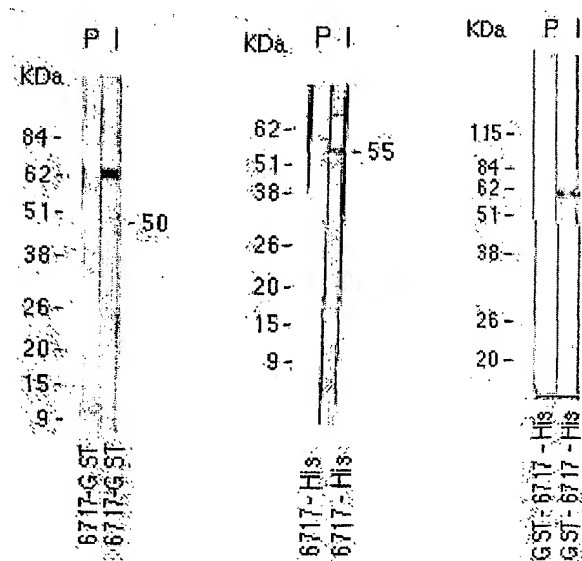


FIG. 38B



39/169

FIGURE 39

FIG. 39A



FIG. 39B

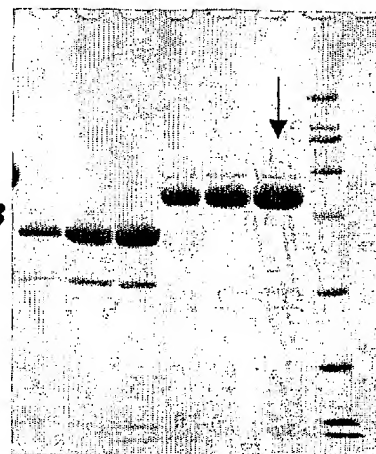


FIG. 39C

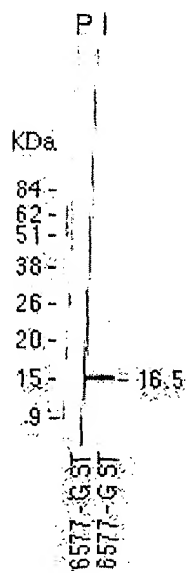
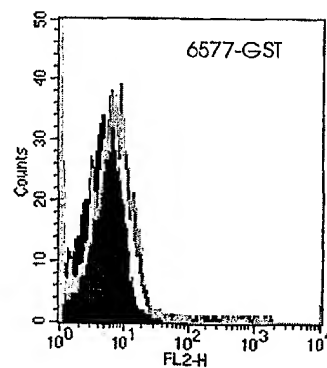


FIG. 39D



40/169

FIGURE 40

Fig. 40A

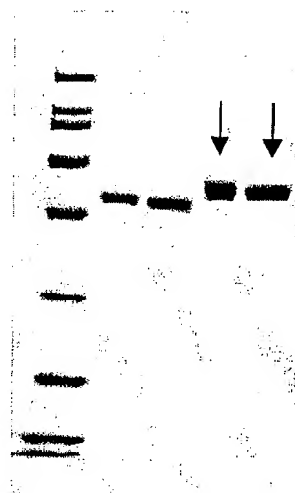
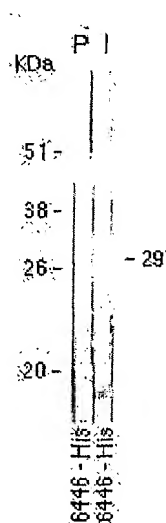


Fig. 40B



41/169

FIGURE 41

Fig. 41A

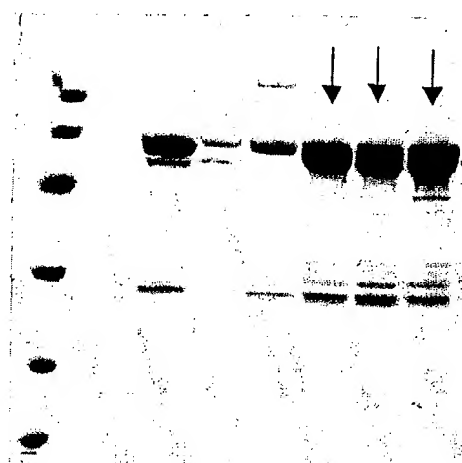


FIG. 41B

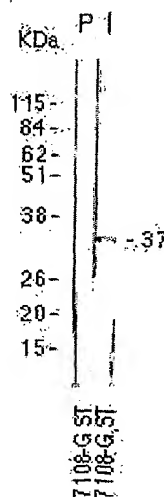
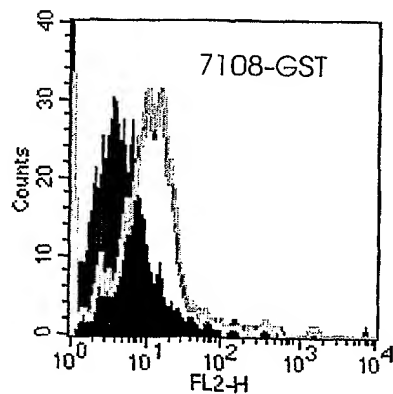


Fig. 41C



42/169

FIGURE 42

FIG. 42A

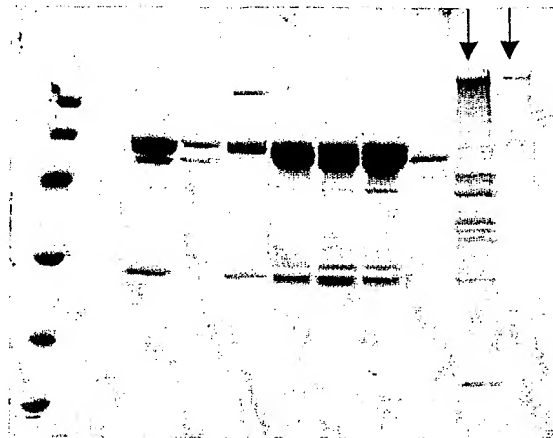


FIG. 42B

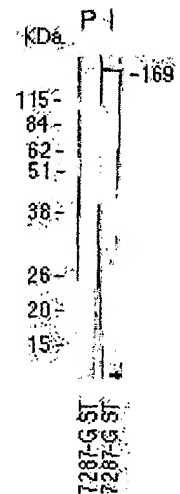
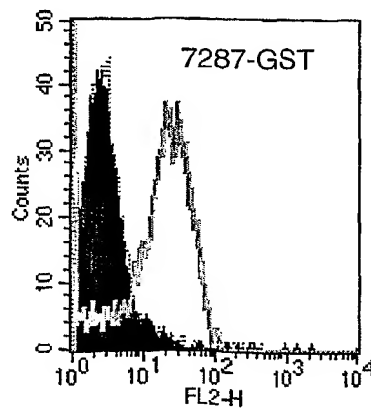


FIG. 42C



43/169

FIGURE 43

FIG. 43A



FIG. 43B

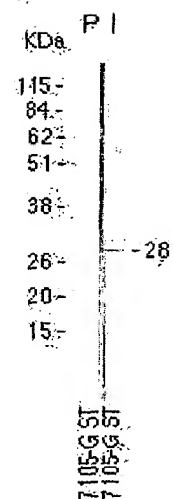
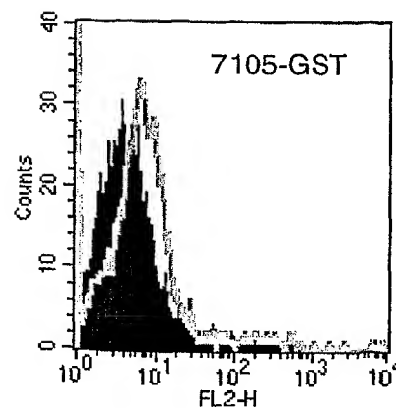


FIG. 43C



44/169

FIGURE 44

FIG. 44A

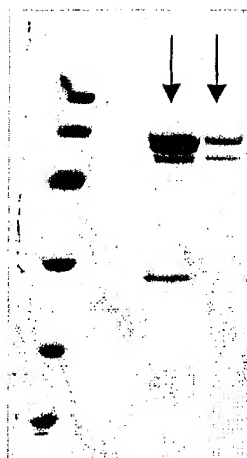


FIG. 44B

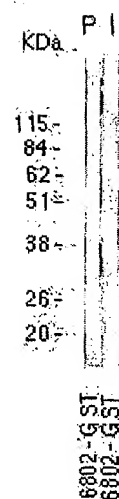
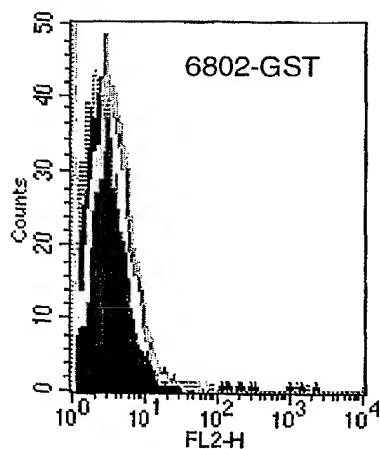


FIG. 44C



45/169

FIGURE 45

FIG. 45A

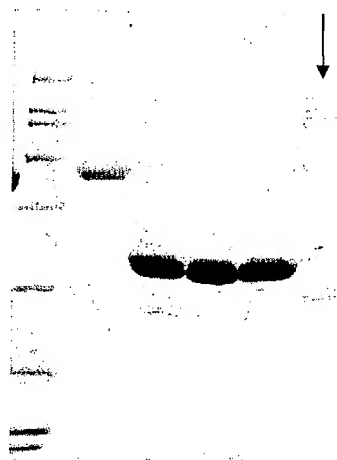


FIG. 45B

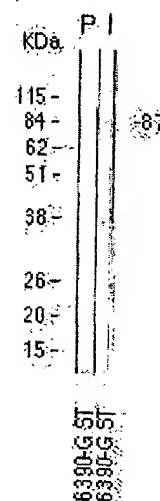
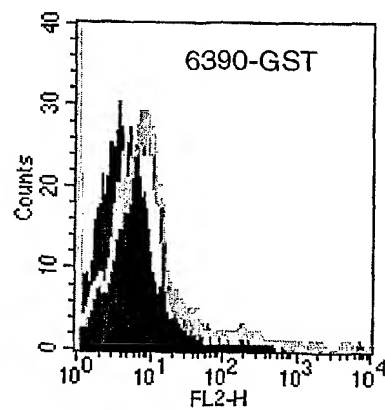


FIG. 45C



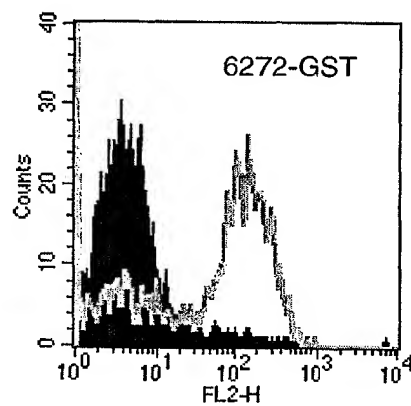
46/169

FIGURE 46

FIG. 46A



FIG. 46B



47/169

FIGURE 47

Fig. 47A

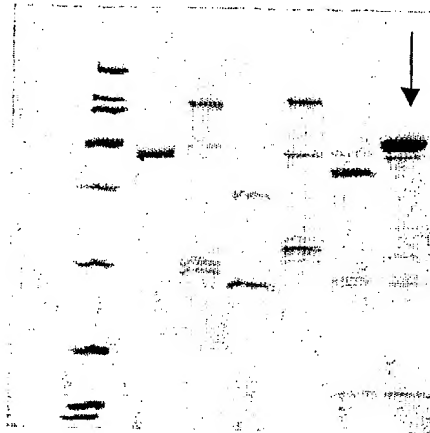


Fig. 47B

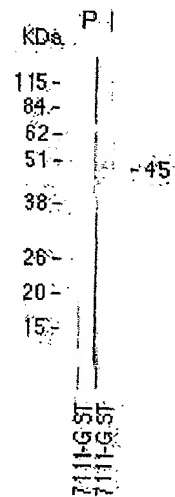
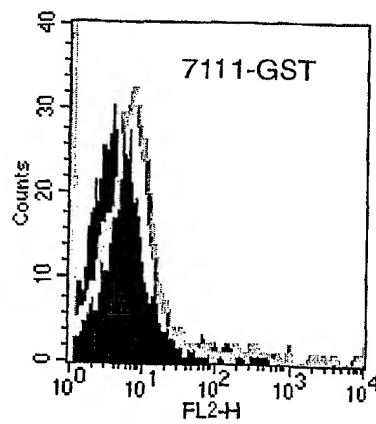


Fig. 47C



48/169

FIGURE 48

Fig. 48A

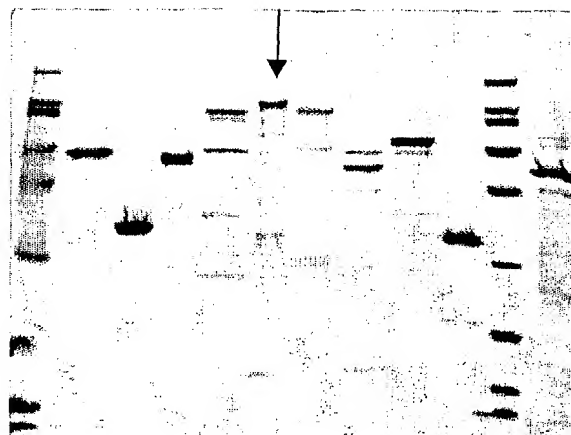


FIG. 48B

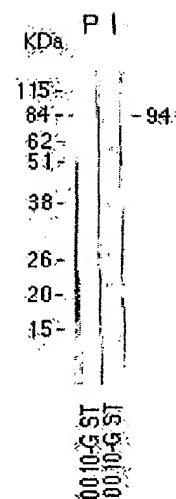
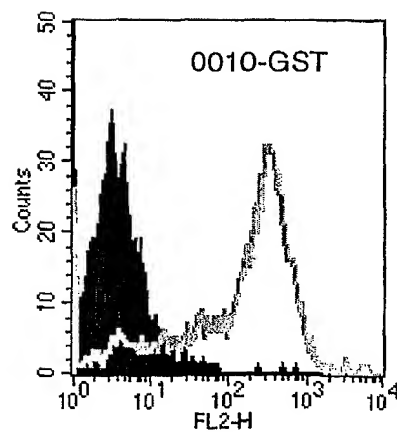


Fig. 48C



49/169

FIGURE 49

Fig. 49A



Fig. 49B

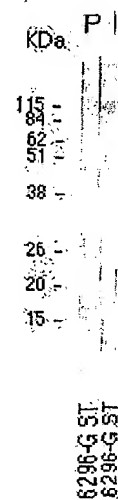
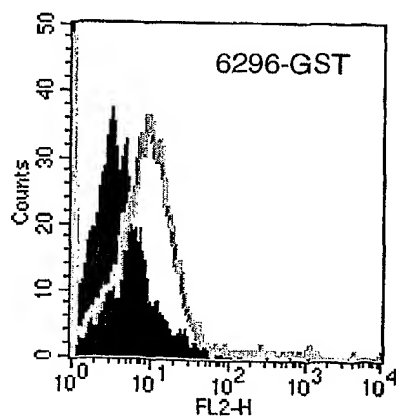


Fig. 49C



50/169

FIGURE 50

Fig. 50A

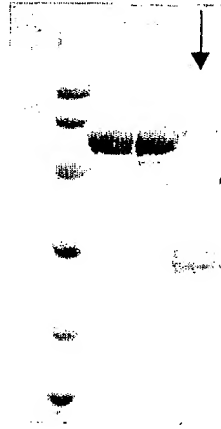


Fig. 50B

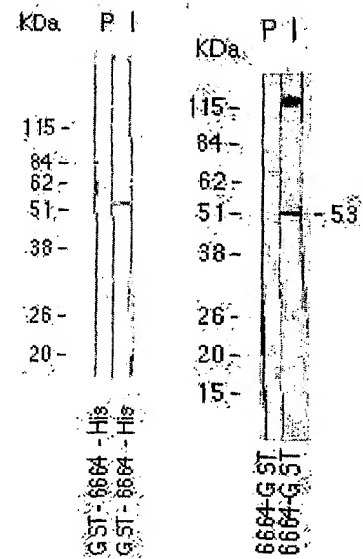
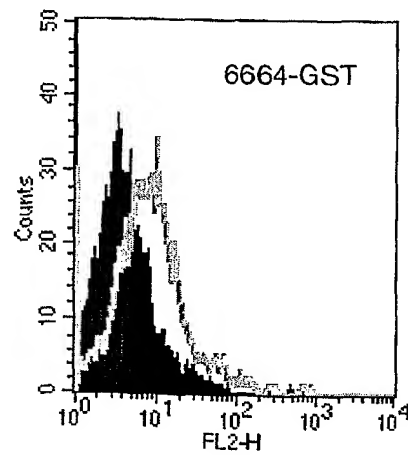


Fig. 50C



51/169

FIGURE 51

FIG. 51A

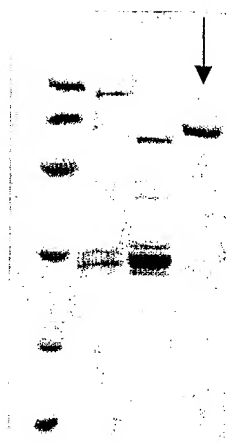


FIG. 51B

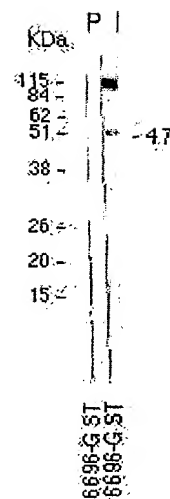
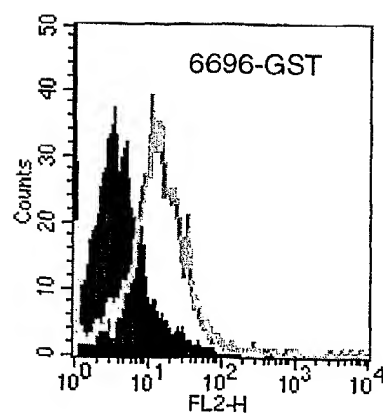


FIG. 51C



52/169

FIGURE 52

FIG. 52A

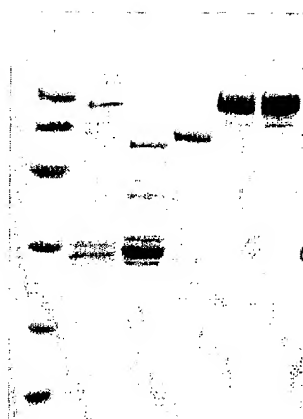


FIG. 52B

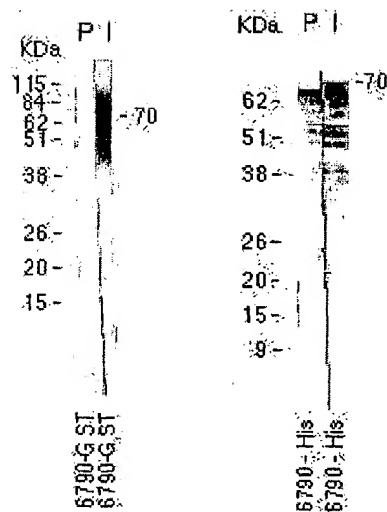
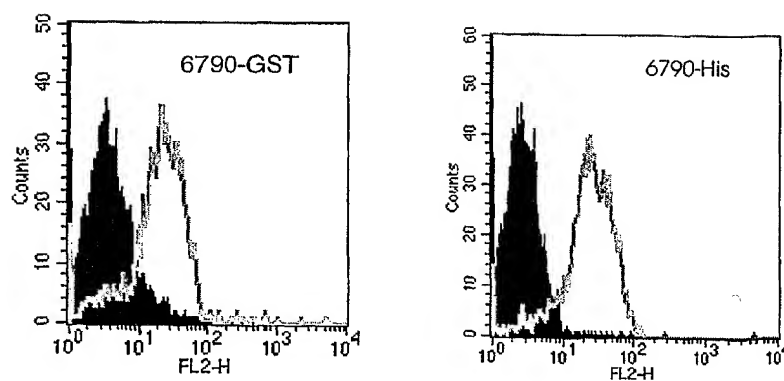


FIG. 52C



53/169

FIGURE 53

FIG. 53A

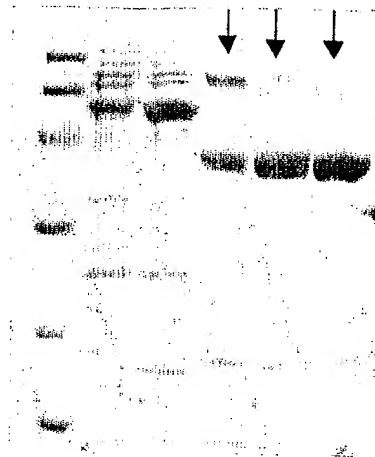
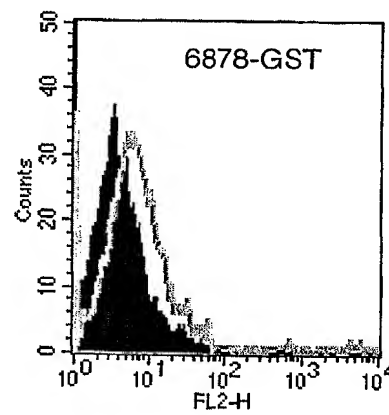


FIG. 53B



54/169

FIGURE 54

Fig. 54A

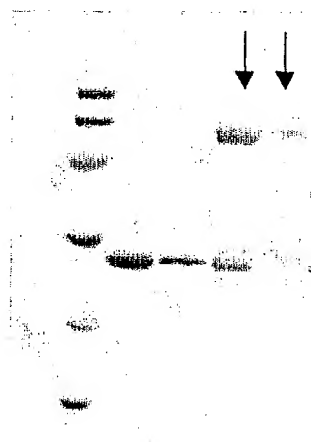


Fig. 54B

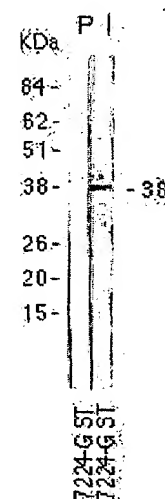
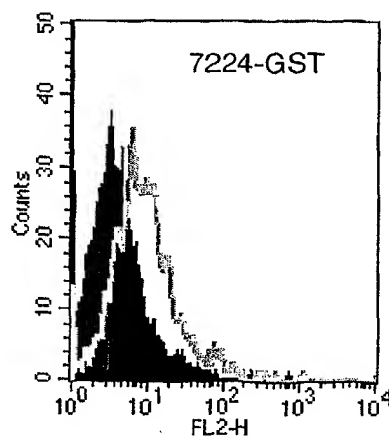


Fig. 54C



55/169

FIGURE 55

Fig. 55A

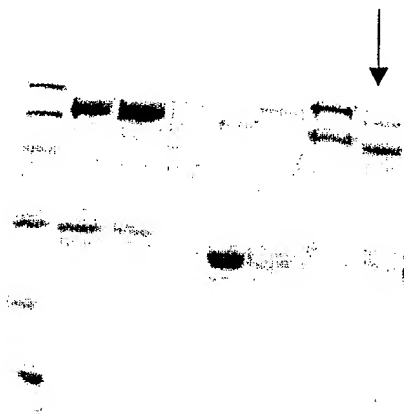


Fig. 55B

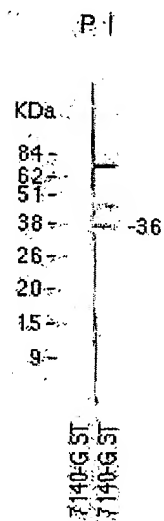
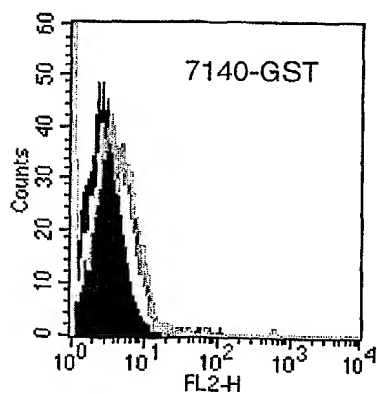


Fig. 55C



56/169

FIGURE 56

FIG. 56A

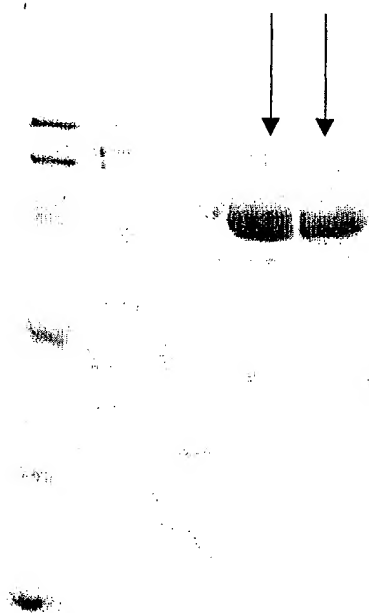


FIG. 56B

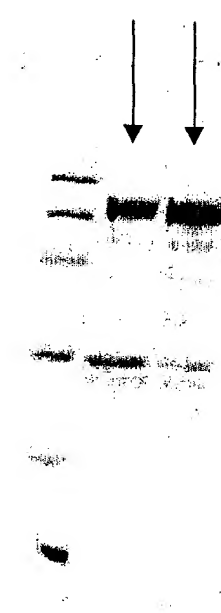


FIG. 56C

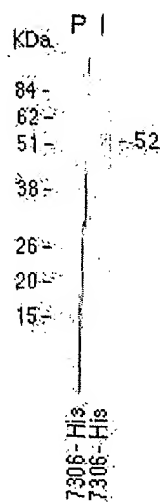
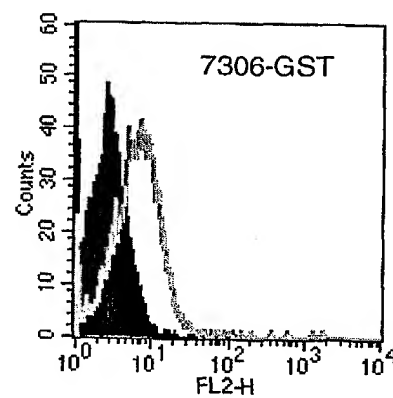


FIG. 56D



57/169

FIGURE 57

FIG. 57A

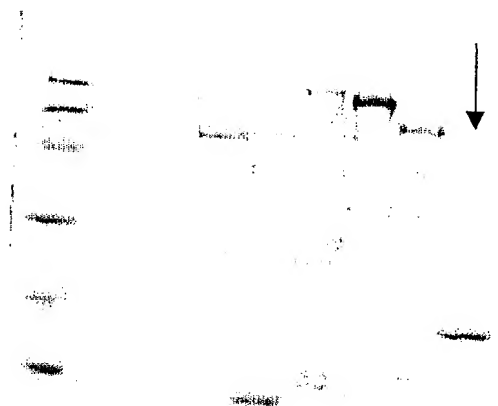


FIG. 57B

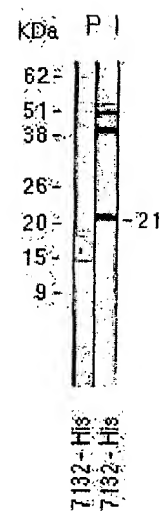
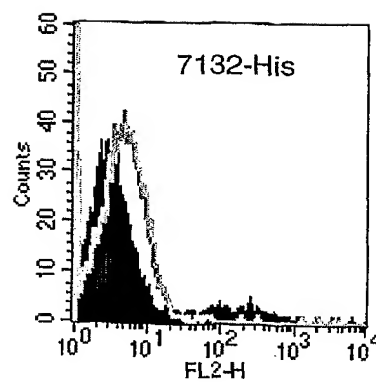


FIG. 57C



58/169

FIGURE 58

FIG. 58A

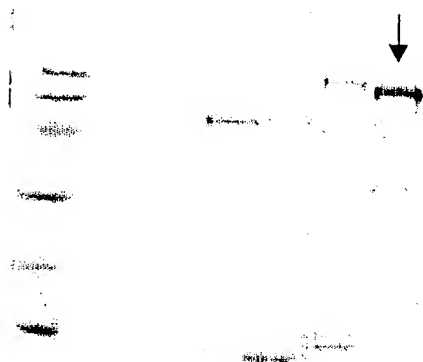


FIG. 58B

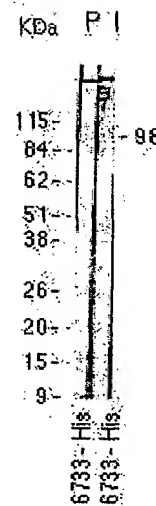
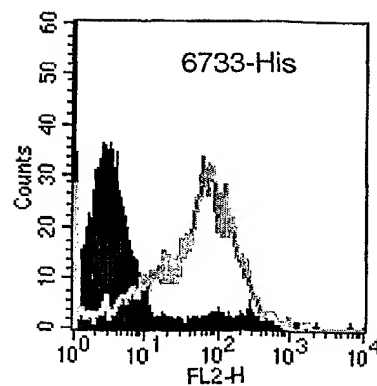


FIG. 58C



59/169

FIGURE 59

Fig. 59A



Fig. 59B

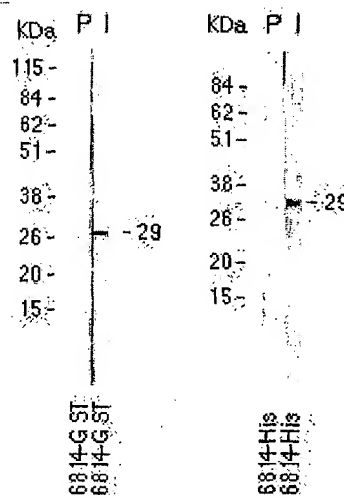
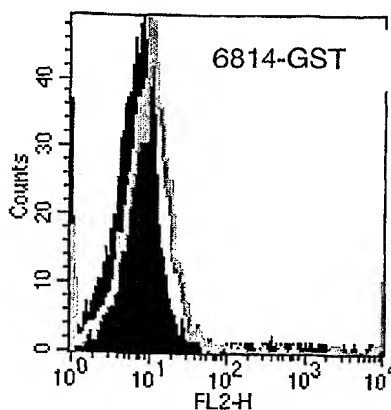


Fig. 59C



60/169

FIGURE 60

FIG. 60A

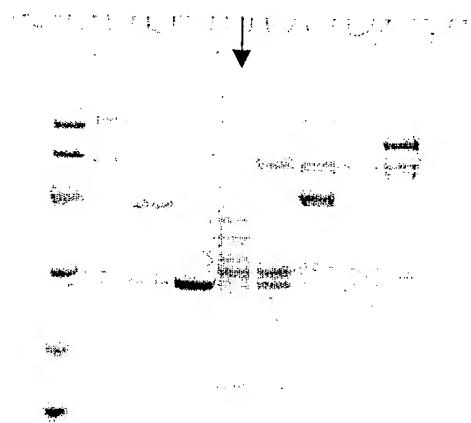


FIG. 60B

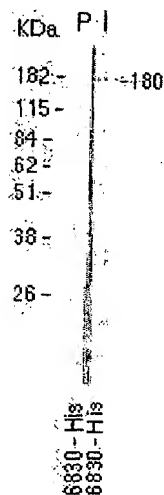
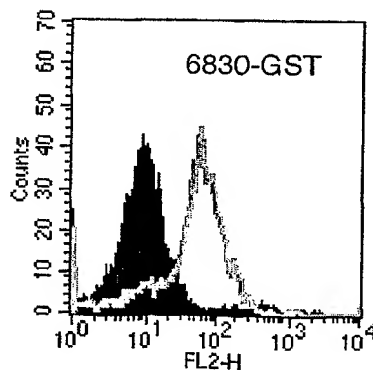


FIG. 60C



61/169

FIGURE 61

Fig. 61A



Fig. 61B

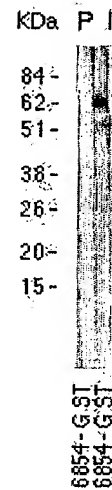
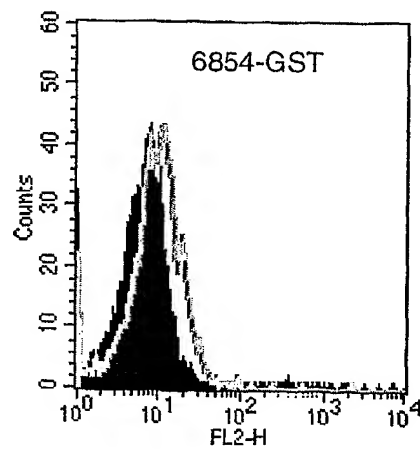


Fig. 61C



62/169

FIGURE 62

Fig. 62A

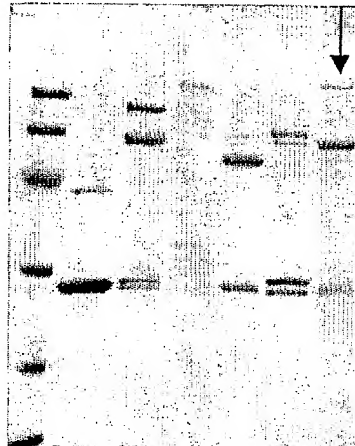


Fig. 62C

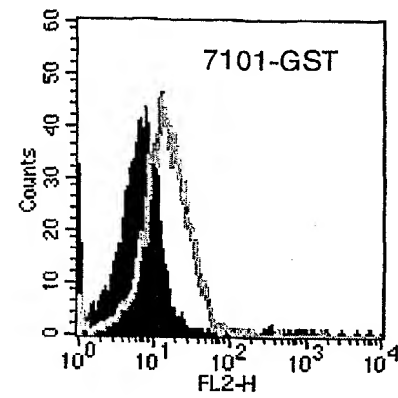
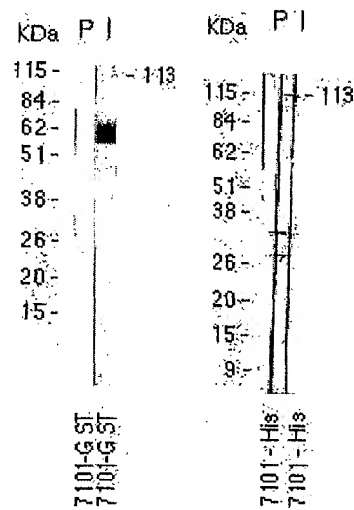


Fig. 62B



63/169

FIGURE 63

Fig. 63A

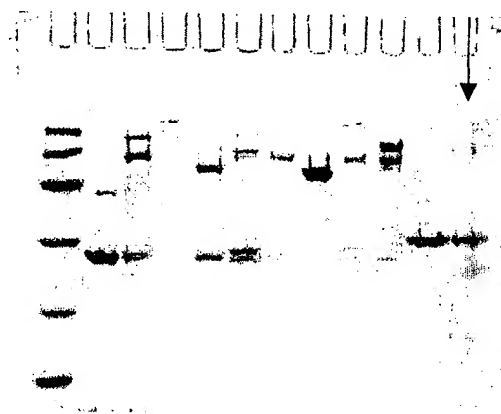


Fig. 63B

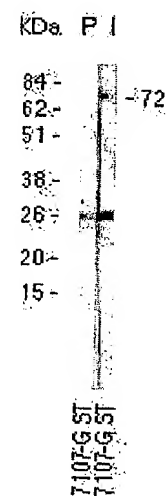
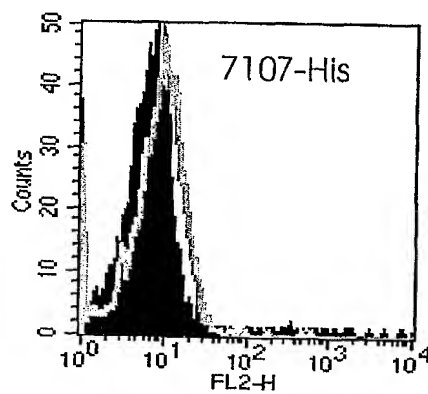


Fig. 63C



64/169

FIGURE 64

FIG. 64A

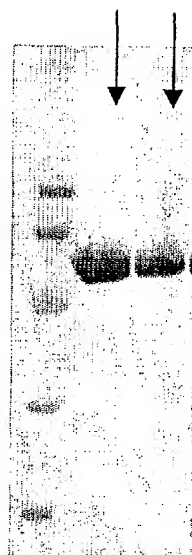


FIG. 64B

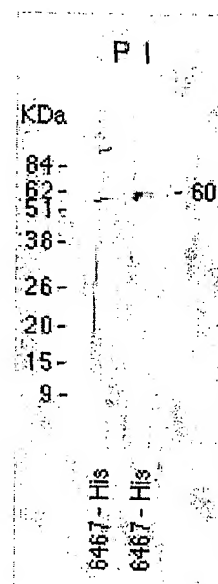


FIG. 64C

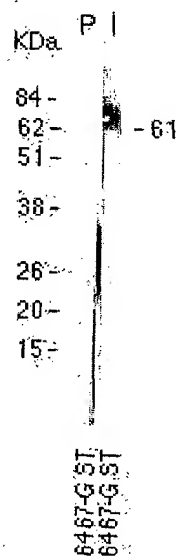
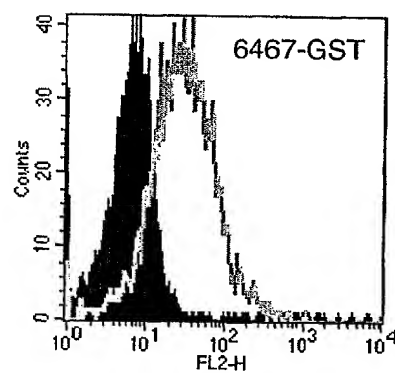


FIG. 64D



65/169

FIGURE 65

Fig. 65A

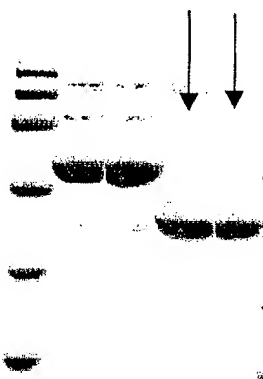


Fig. 65B

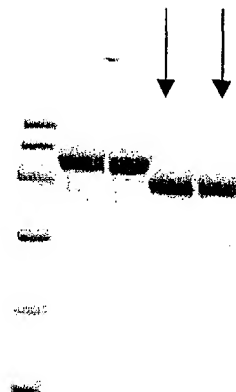
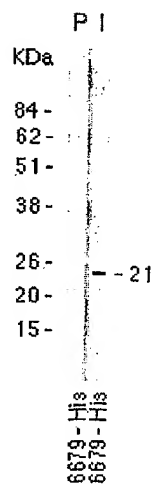


Fig. 65C



66/169

FIGURE 66

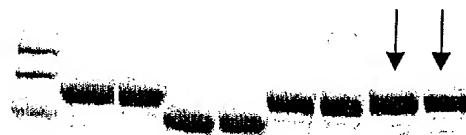


FIG. 66A

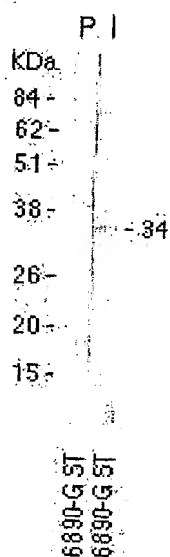


FIG. 66B

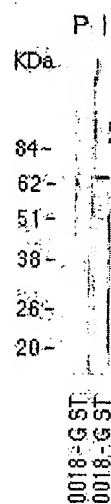
67/169

FIGURE 67

Fig. 67A



Fig. 67B



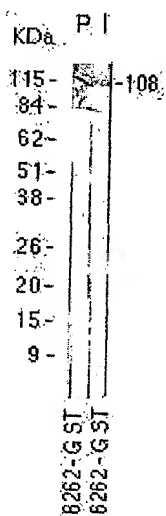
68/169

FIGURE 68

Fig. 68A



Fig. 68B



69/169

FIGURE 69

Fig. 69A

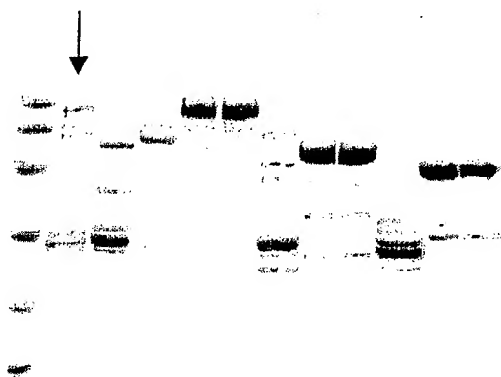
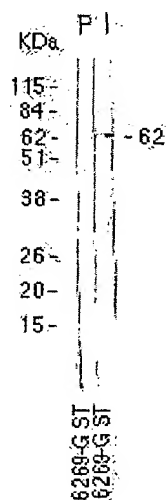


Fig. 69B



70/169

FIGURE 70

Fig. 70A

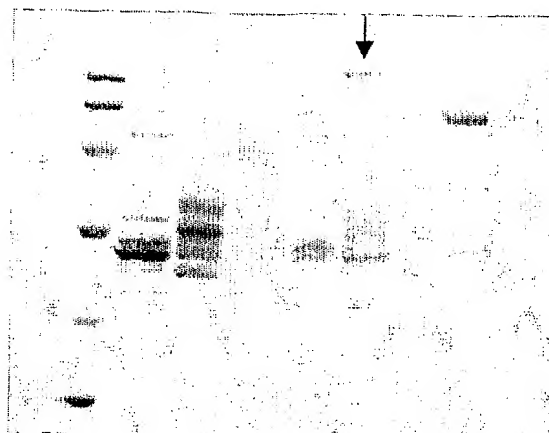
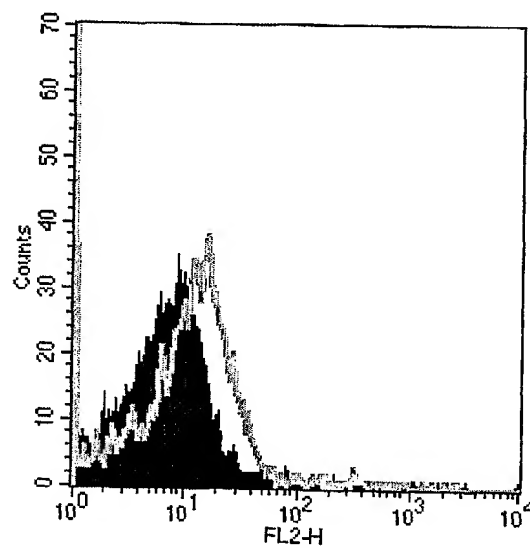


Fig. 70B



71/169

FIGURE 71

FIG. 71A

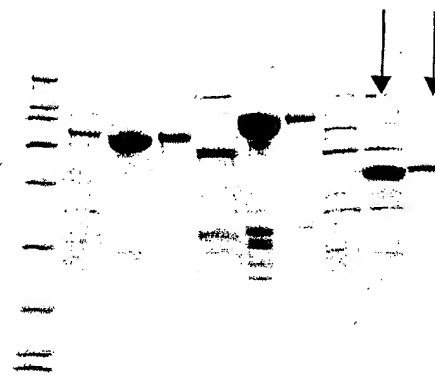
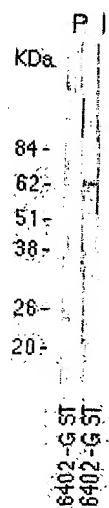


FIG. 71B



72/169

FIGURE 72

Fig. 72A

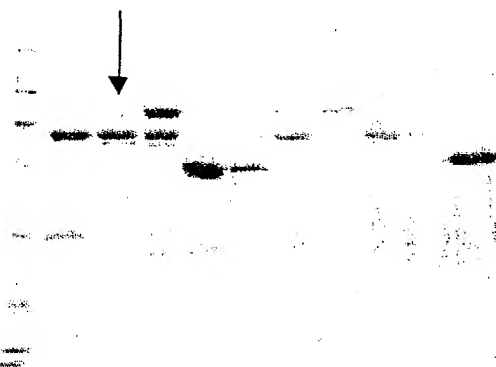
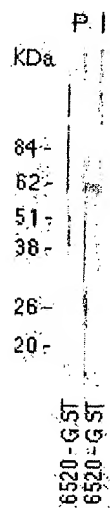


Fig. 72B



73/169

FIGURE 73

Fig. 73A

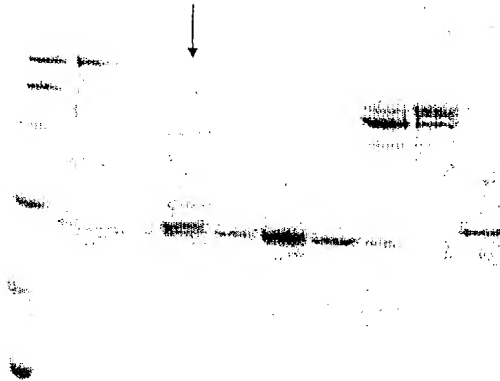
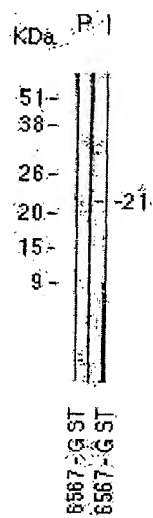


Fig. 73B



74/169

FIGURE 74

FIG. 74A

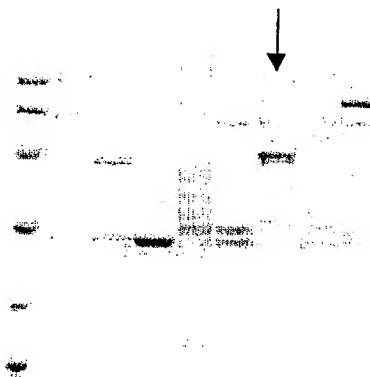


FIG. 74B

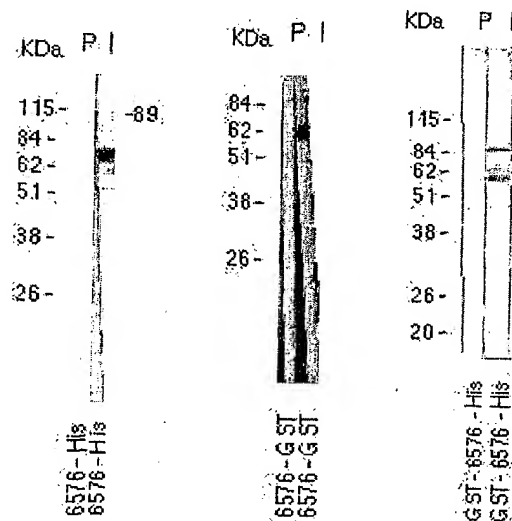
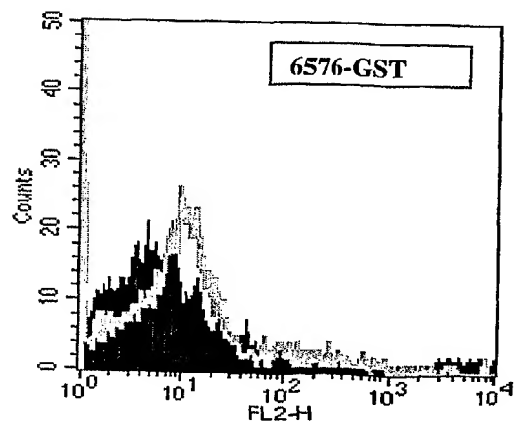


FIG. 74C



75/169

FIGURE 75

FIG. 75A

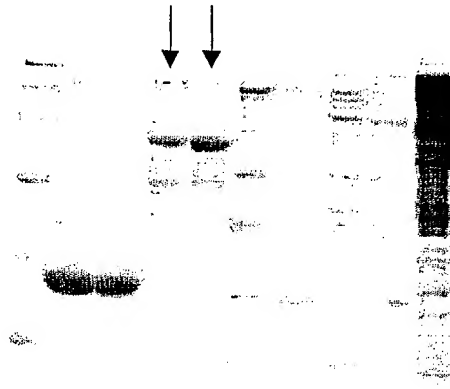
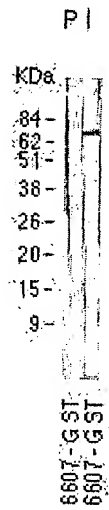


FIG. 75B



76/169

FIGURE 76

Fig. 76A

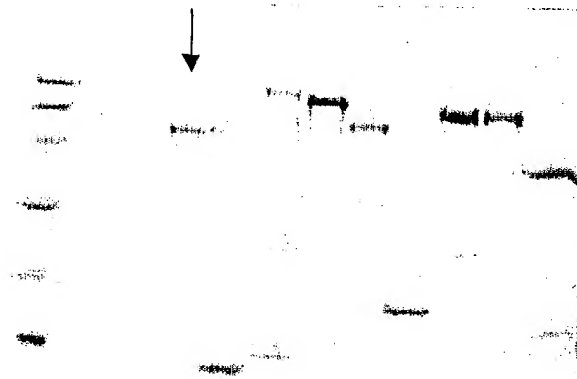
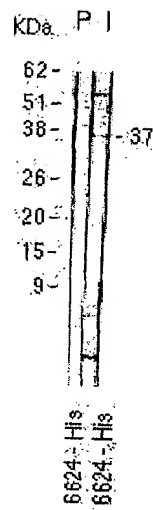


Fig. 76B



77/169

FIGURE 77



Fig. 77A

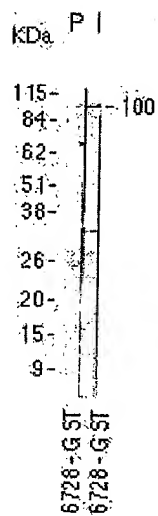


Fig. 77B

78/169

FIGURE 78

Fig. 78A

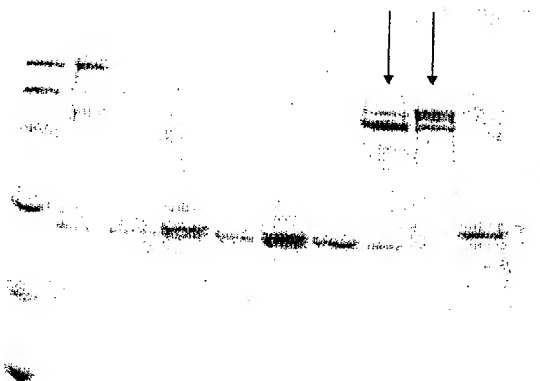
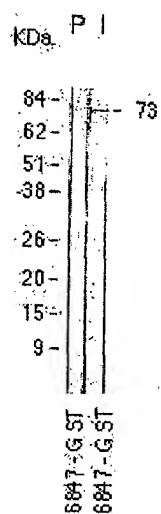


Fig. 78B



79/169

FIGURE 79

Fig. 79A

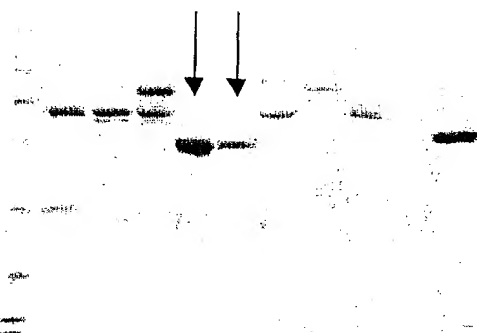
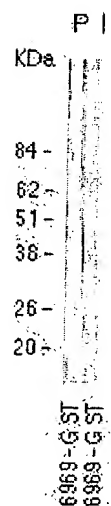


Fig. 79B



80/169

FIGURE 80

Fig. 80A

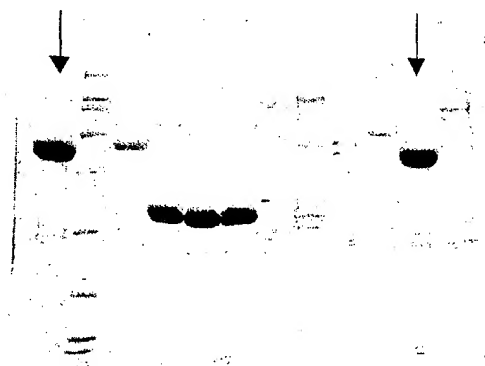
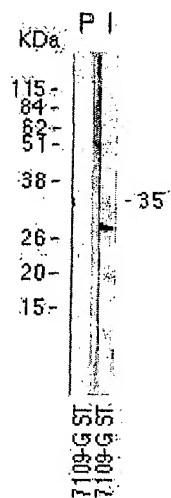


Fig. 80B



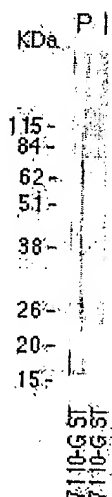
81/169

FIGURE 81

FIG. 81A



FIG. 81B



82/169

FIGURE 82

Fig. 82A

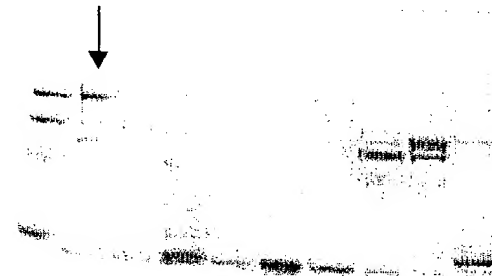
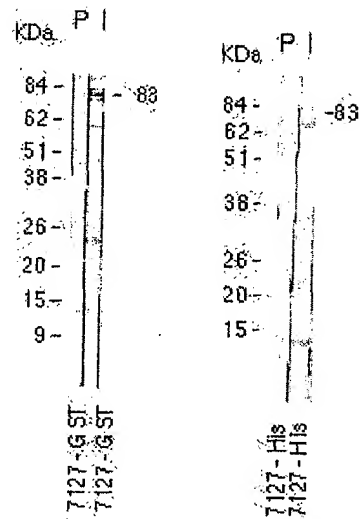


Fig. 82B



83/169

FIGURE 83

Fig. 83A

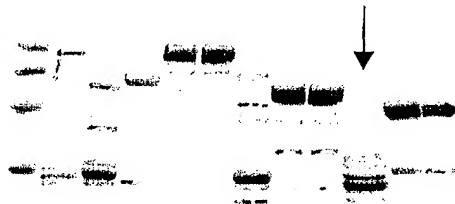
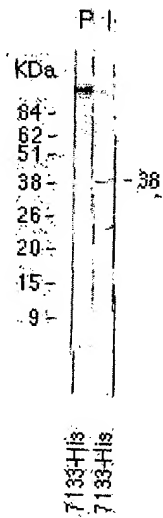


Fig. 83B



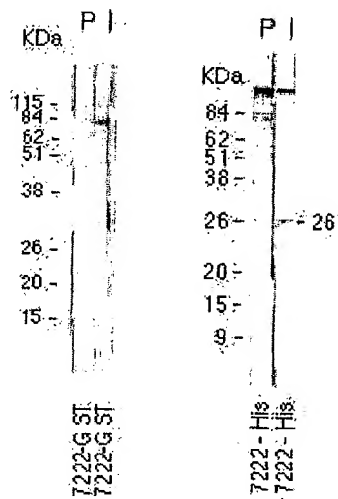
84/169

FIGURE 84

Fig. 84A



Fig. 84B



85/169

FIGURE 85

FIG. 85A

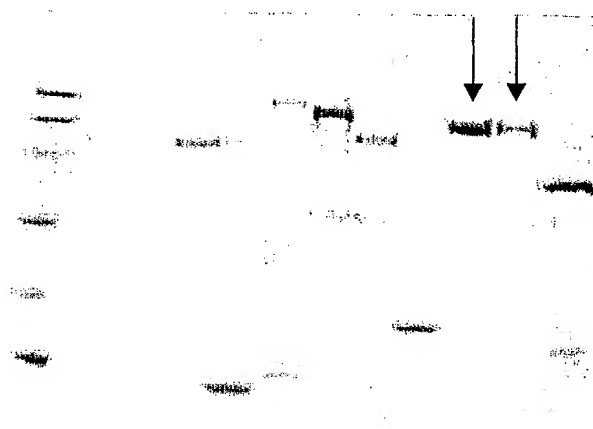
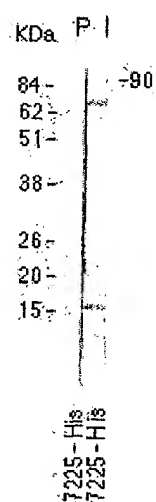


FIG. 85B



86/169

FIGURE 86

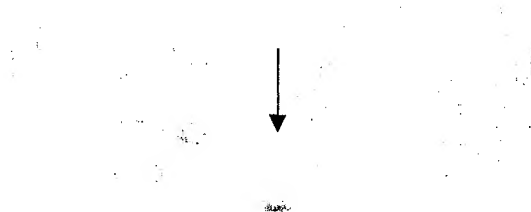


Fig. 86A

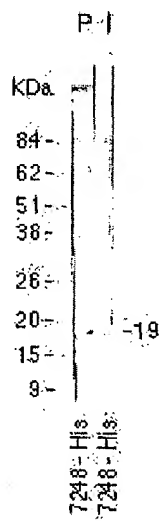


Fig. 86B

87/169

FIGURE 87

Fig. 87A

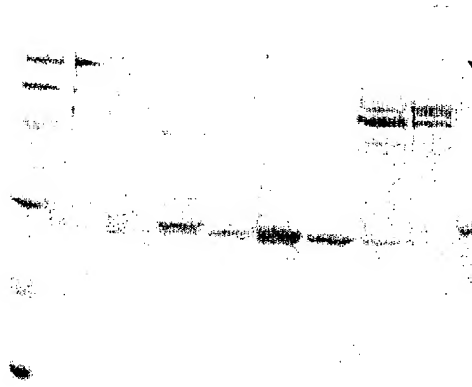
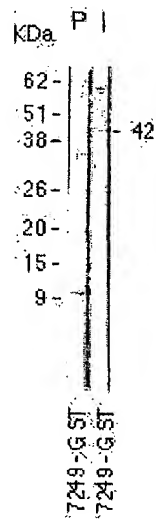


Fig. 87B



88/169

FIGURE 88

FIG. 88A

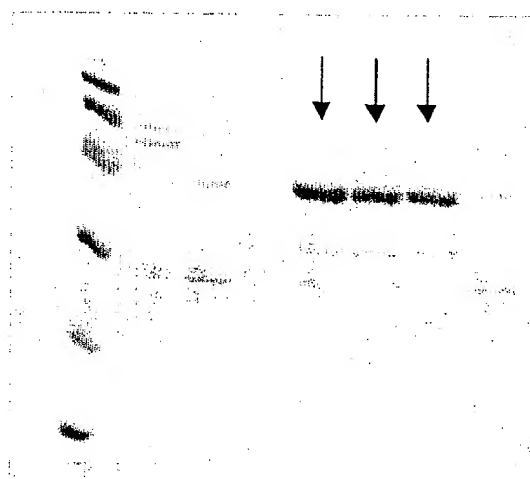
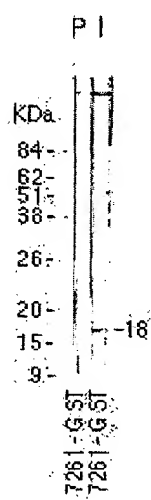
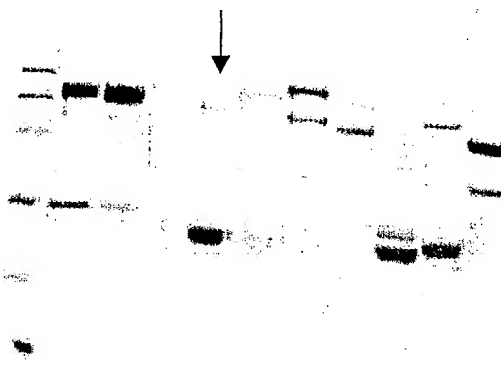
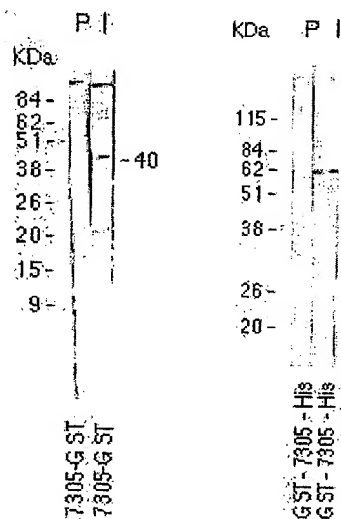


FIG. 88B



89/169

FIGURE 89**Fig. 89A****Fig. 89B**

90/169

FIGURE 90

Fig. 90A

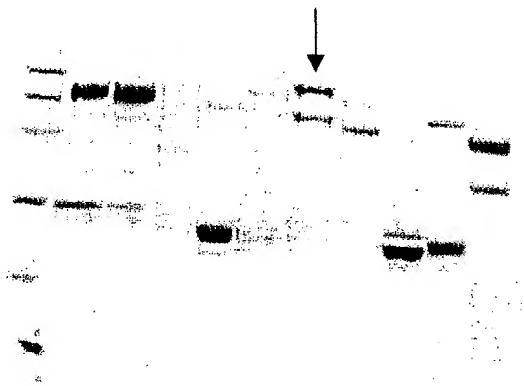
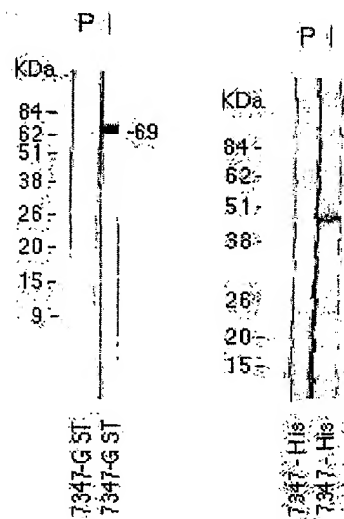


Fig. 90B



91/169

FIGURE 91

Fig. 91A

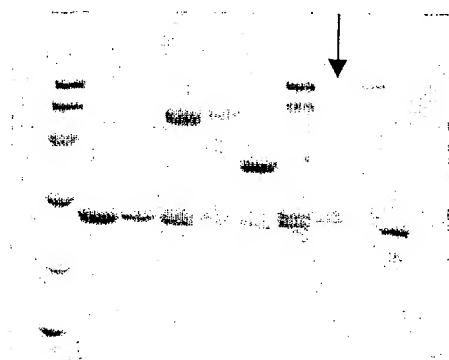
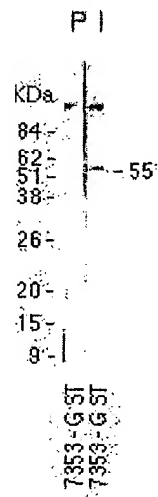


Fig. 91B



92/169

FIGURE 92

FIG. 92A

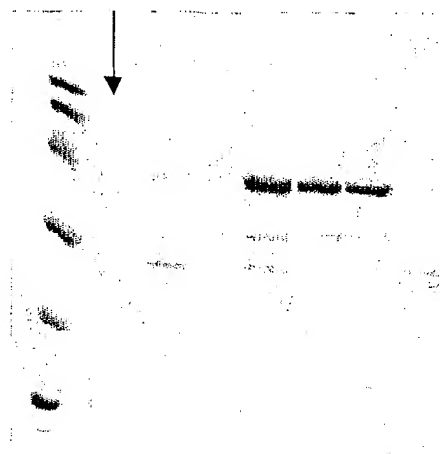
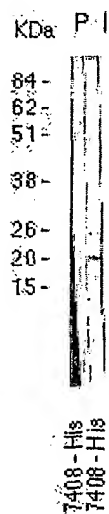


FIG. 92B



93/169

FIGURE 93

Fig. 93A

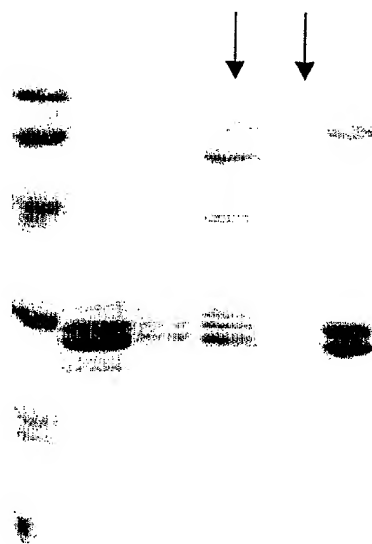


Fig. 93B

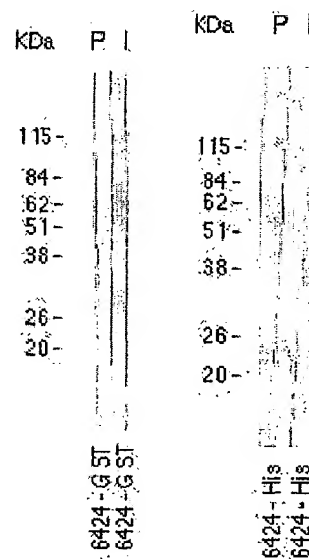
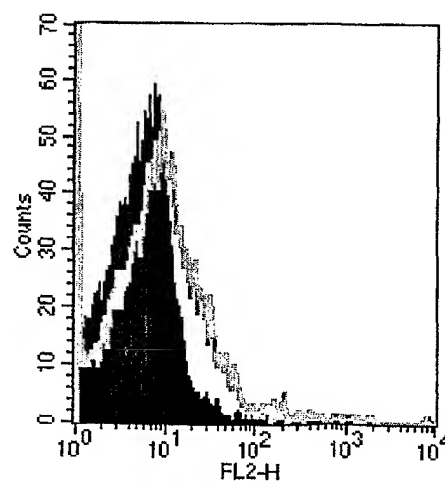


Fig. 93C



94/169

FIGURE 94

FIG. 94A

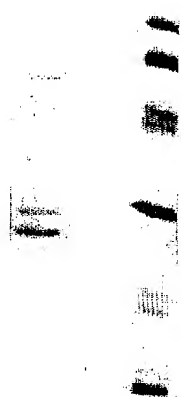


FIG. 94B

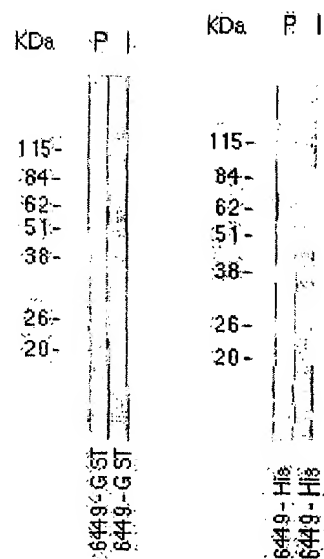
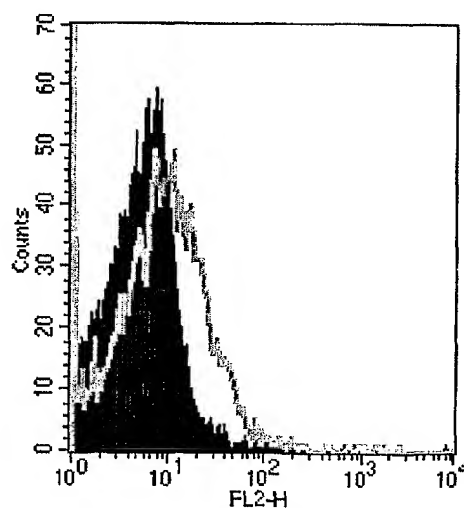


FIG. 94C



95/169

FIGURE 95

Fig. 95A



Fig. 95B

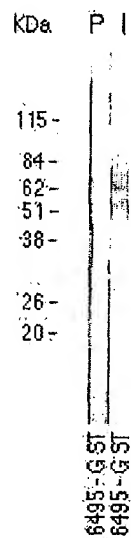
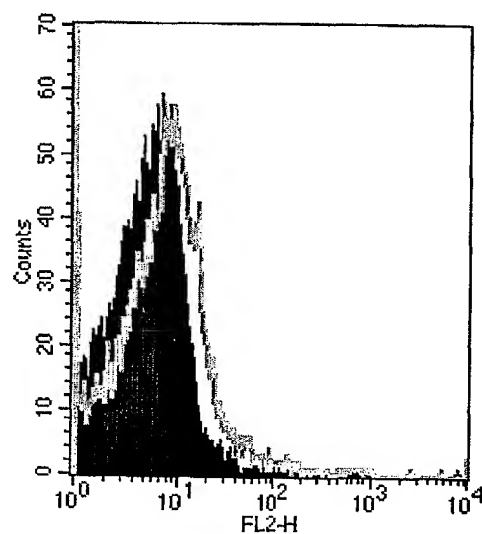


Fig. 95C



96/169

FIGURE 96

FIG. 96A

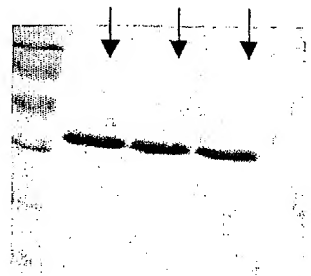


FIG. 96B



FIG. 96C

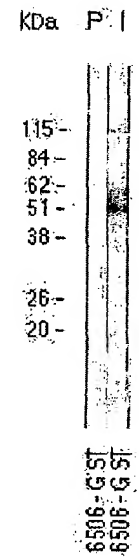
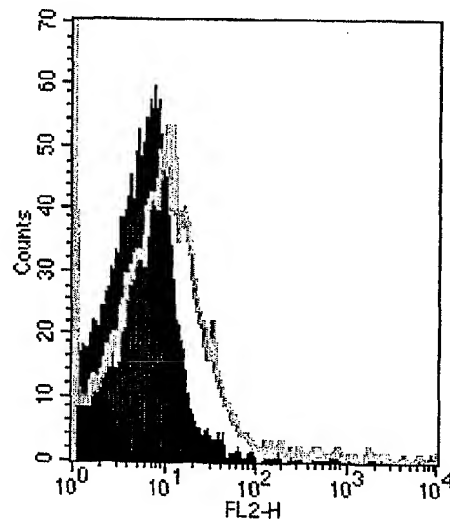


Fig. 96D



97/169

FIGURE 97

Fig. 97A

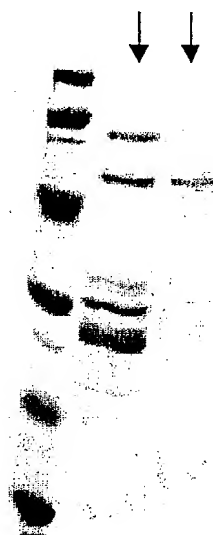


Fig. 97B

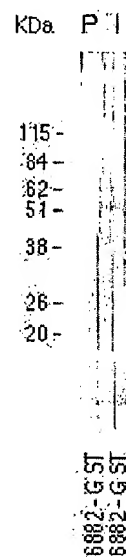
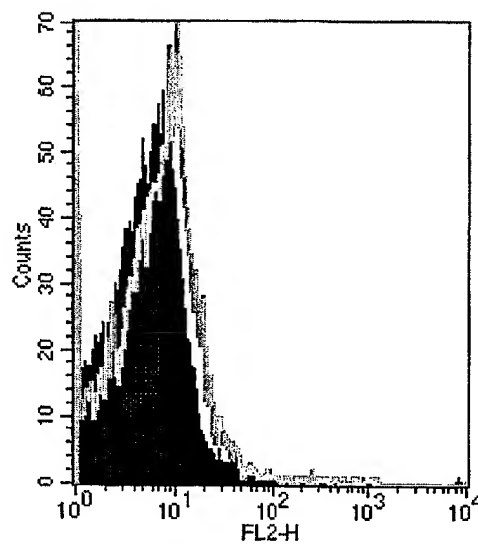


Fig. 97C



98/169

FIGURE 98

FIG. 98A

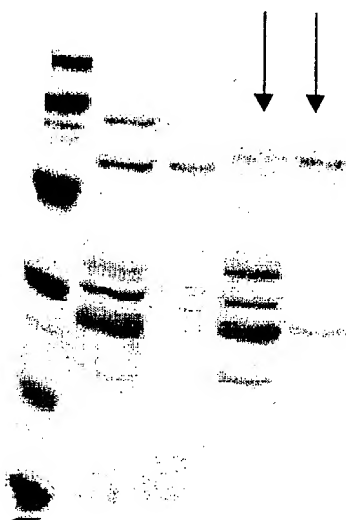


FIG. 98B

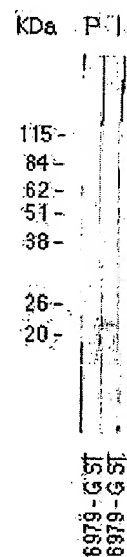
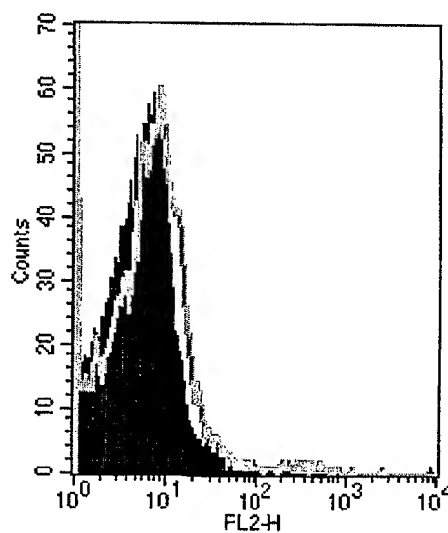


FIG. 98C



99/169

FIGURE 99

FIG. 99A

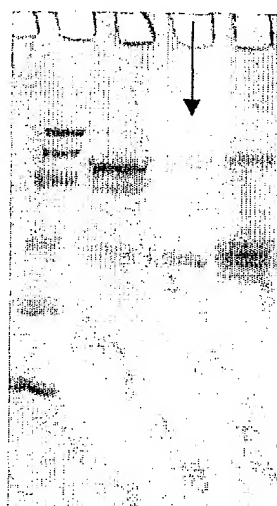


FIG. 99B

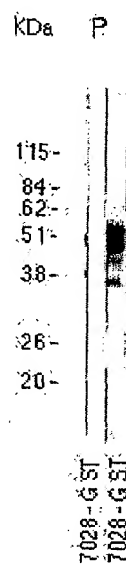
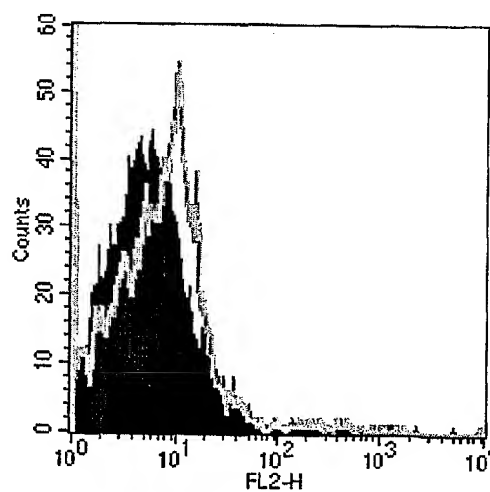


FIG. 99C



100/169

FIGURE 100

Fig. 100A

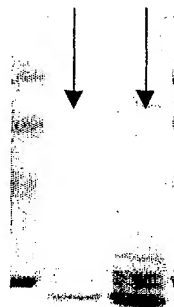


Fig. 100B

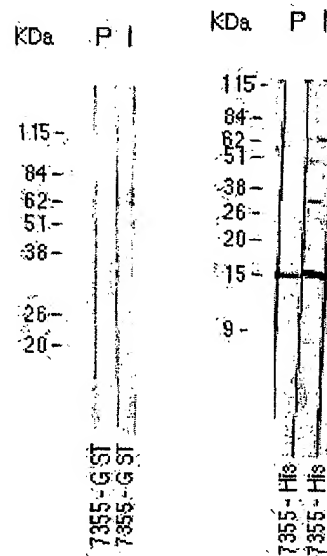
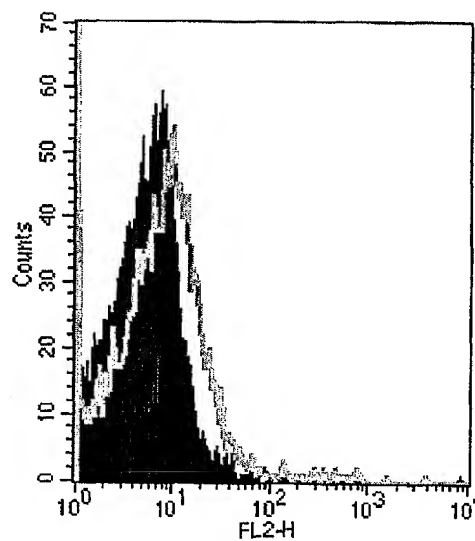


Fig. 100C



101/169

FIGURE 101

FIG. 101A

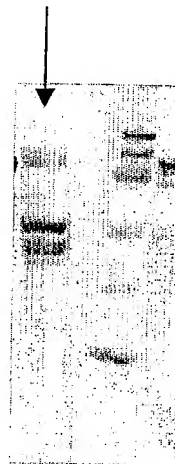


FIG. 101B

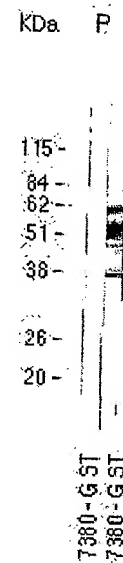
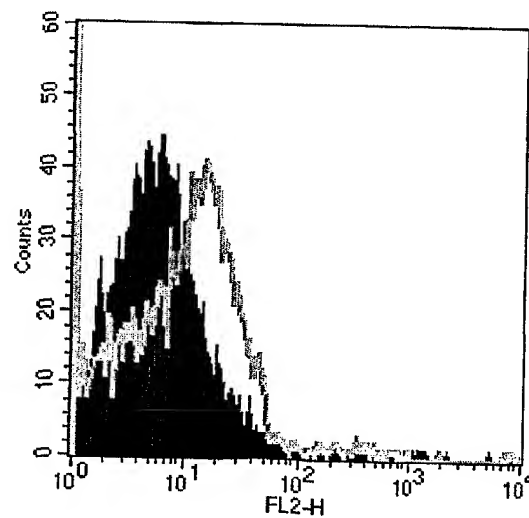


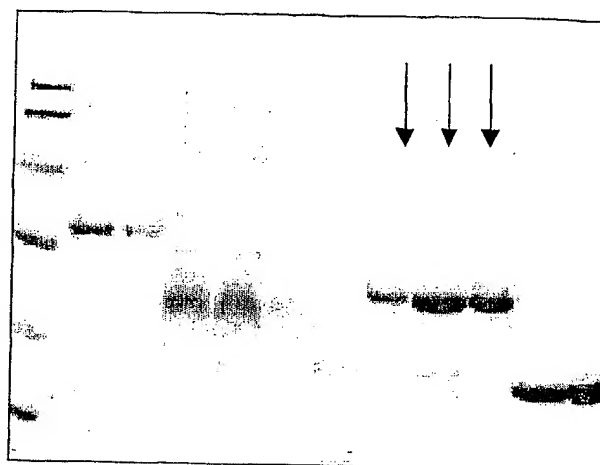
FIG. 101C



102/169

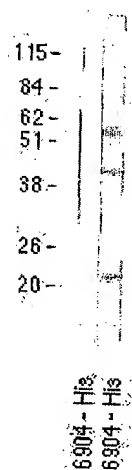
FIGURE 102

FIG. 102A



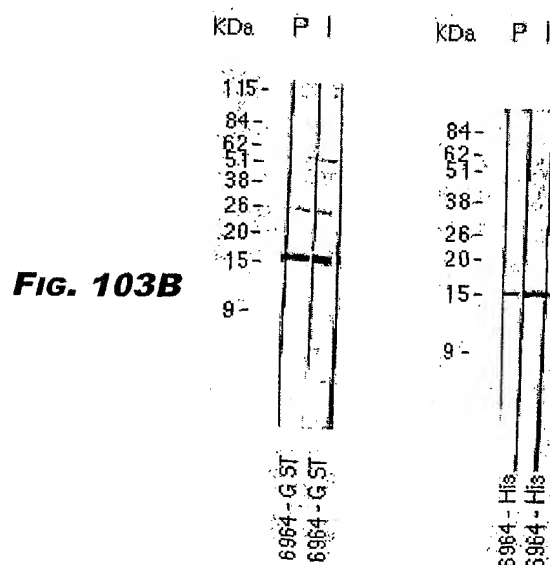
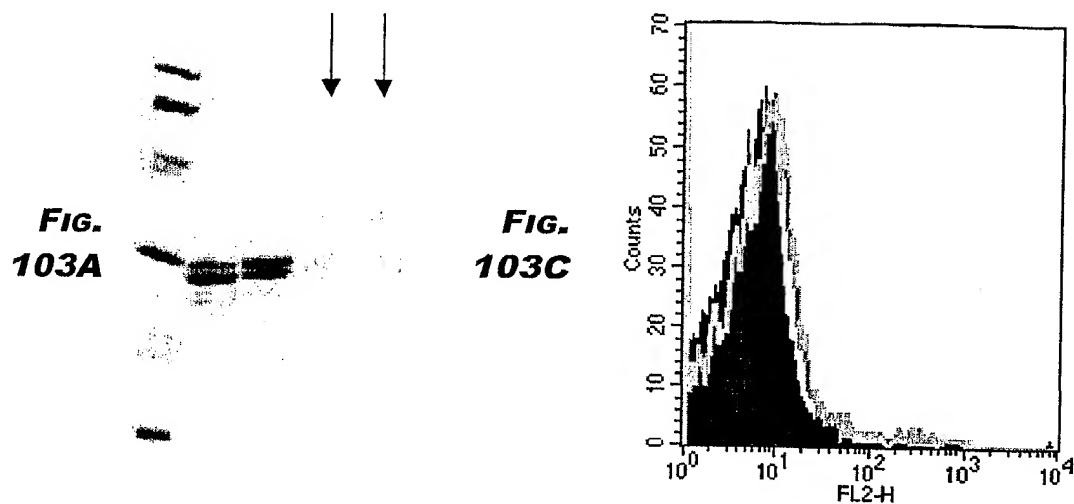
KDa P I

FIG. 102B



103/169

FIGURE 103



104/169

FIGURE 104

FIG. 104A

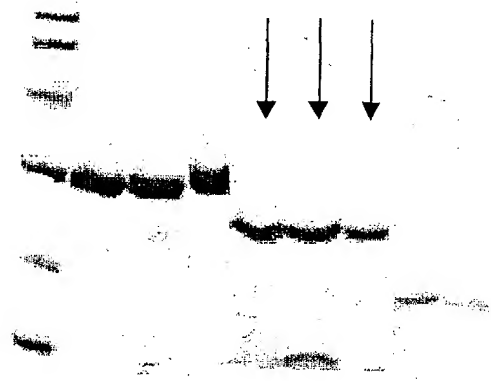
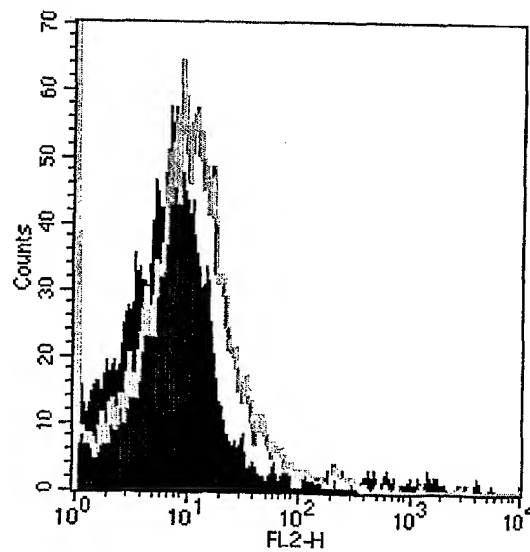


FIG. 104B



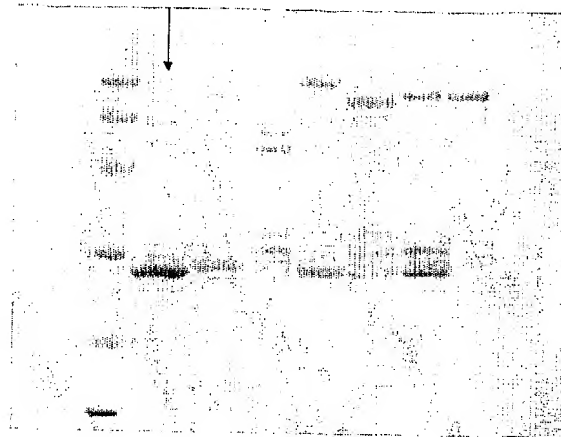
FIG. 104C



105/169

FIGURE 105

Fig. 105A



kDa P I

115-
84-
62-
51-
38-
26-
20-

Fig. 105B

6281-G ST
6281-G ST

106/169

FIGURE 106

Fig. 106A

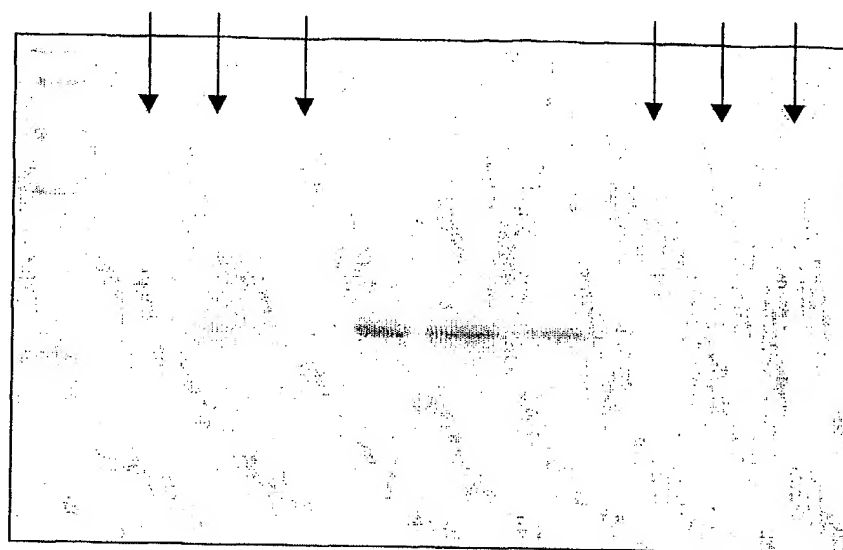
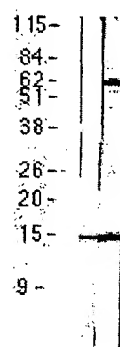


Fig. 106B

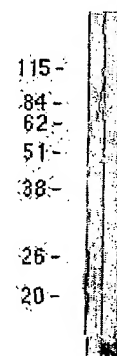
kDa P I



63016 - His
63016 - His

FIGURE 107

kDa P I



6434 - His
6434 - His

107/169

FIGURE 108

Fig. 108A

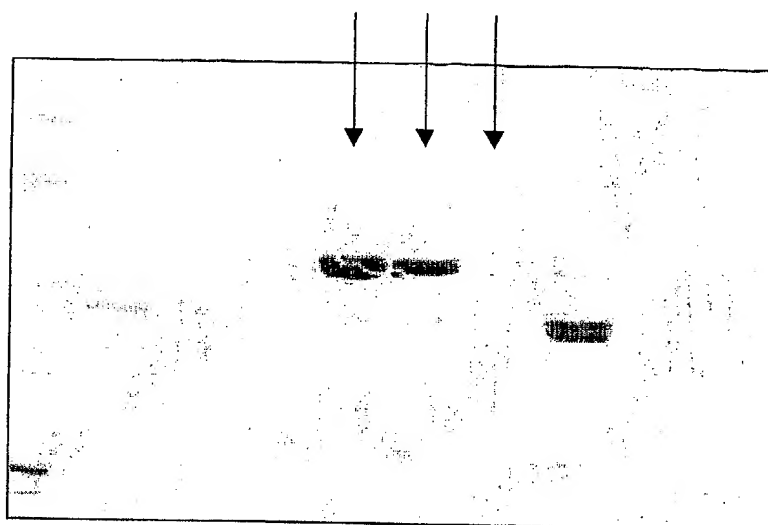
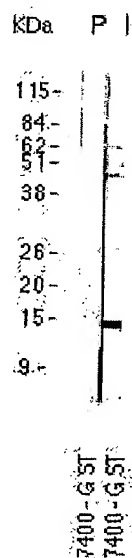


Fig. 108B



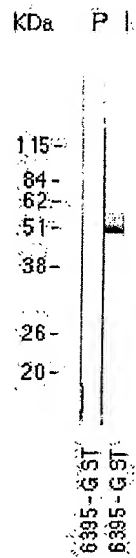
108/169

FIGURE 109

Fig. 109A



Fig. 109B



109/169

FIGURE 110



FIG. 110A

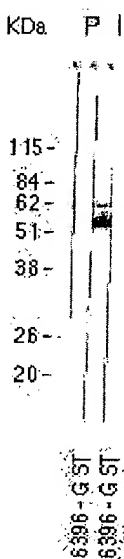


FIG. 110B

110/169

FIGURE 111



FIG. 111A

KDa P I

115-
84-
62-
51-
38-
26-
20-

FIG. 111B

6408 - His
6408 - His

111/169

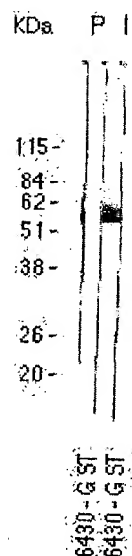
FIGURE 112



FIG. 112A



FIG. 112B



112/169

FIGURE 113



FIG. 113A

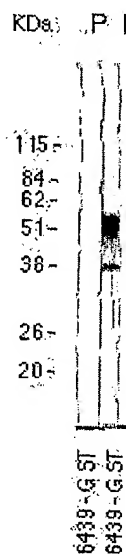


FIG. 113B

113/169

FIGURE 114

FIG. 114A

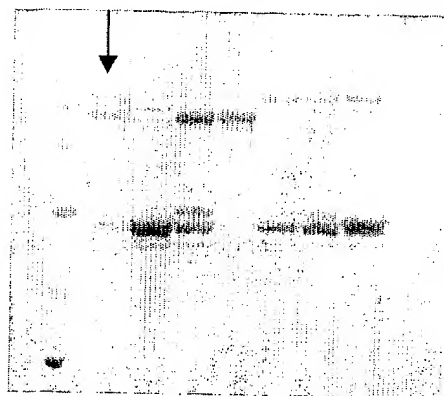
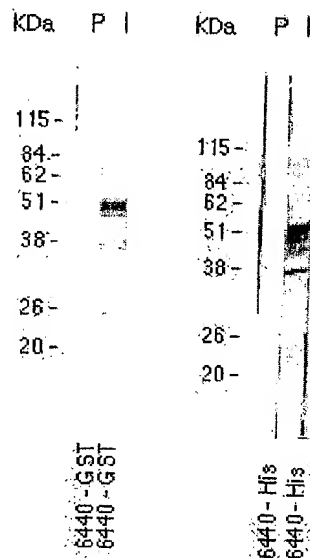


FIG. 114B



114/169

FIGURE 115

FIG. 115A

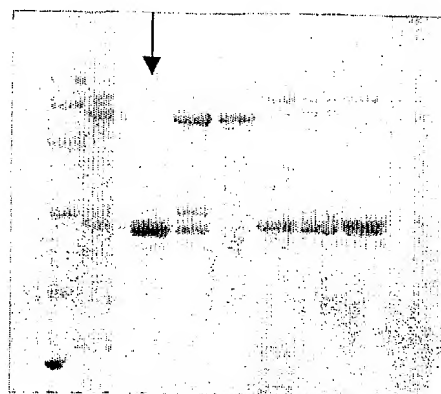
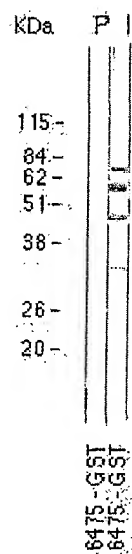


FIG. 115B



115/169

FIGURE 116

Fig. 116A

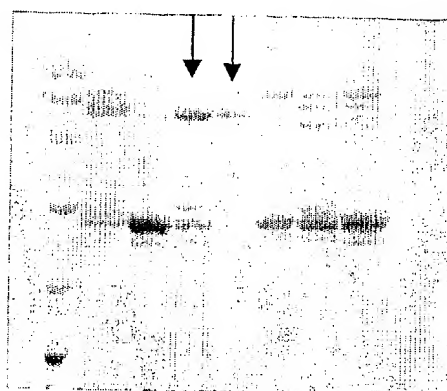
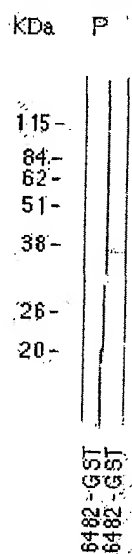


Fig. 116B



116/169

FIGURE 117

FIG. 117A

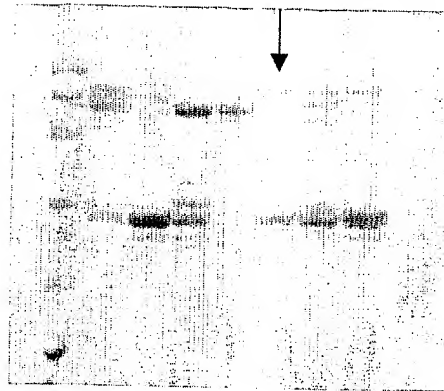
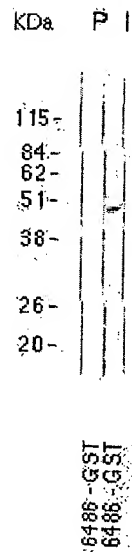


FIG. 117B



117/169

FIGURE 118

Fig. 118A

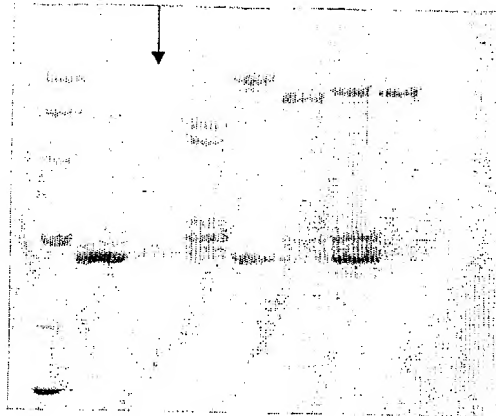
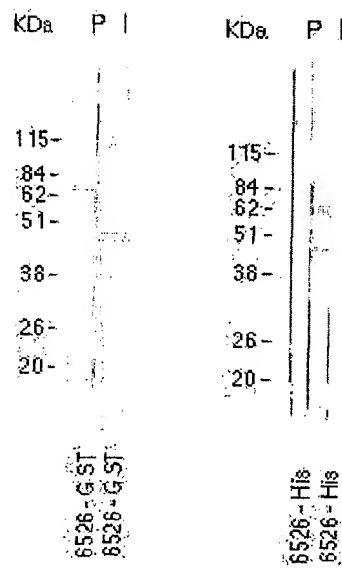


Fig. 118B



118/169

FIGURE 119

Fig. 119A

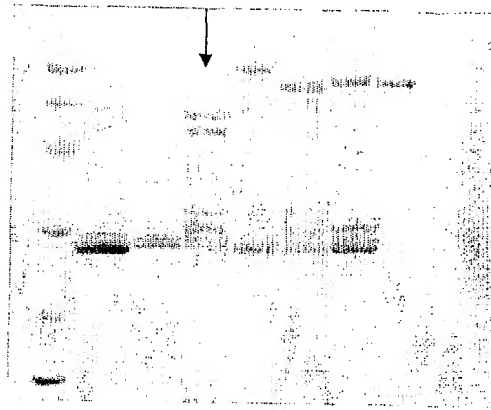
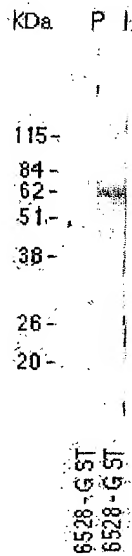


Fig. 119B



119/169

FIGURE 120

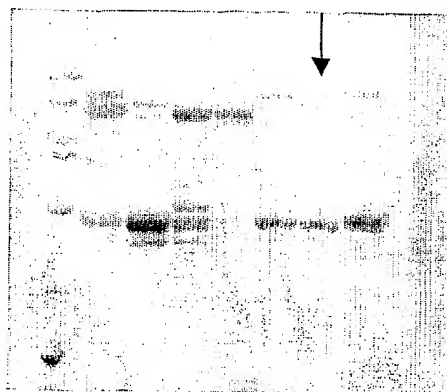


FIG. 120A

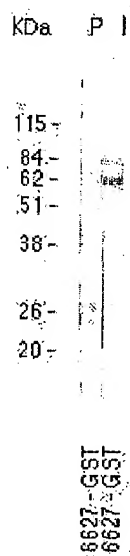


FIG. 120B

120/169

FIGURE 121

FIG. 121A

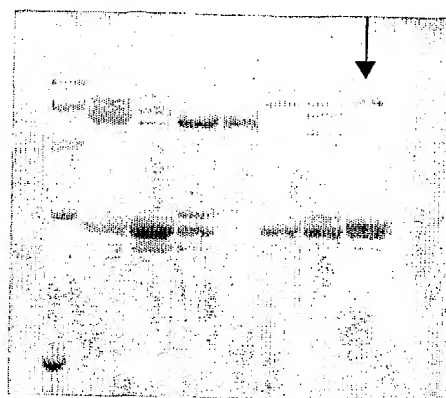
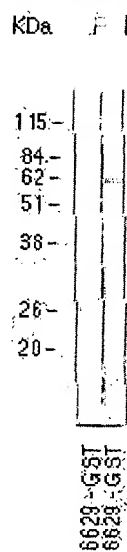


FIG. 121B



121/169

FIGURE 122

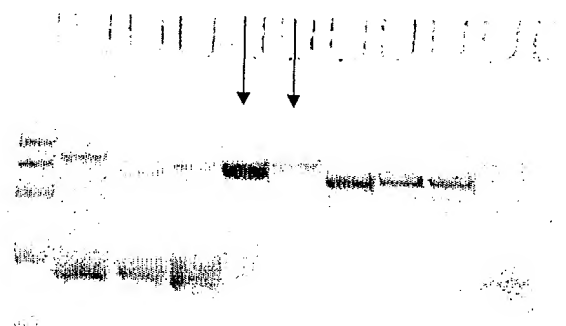


FIG. 122A

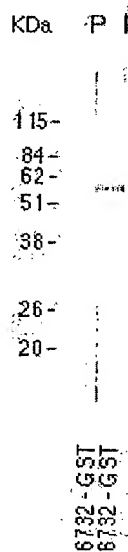


FIG. 122B

122/169

FIGURE 123

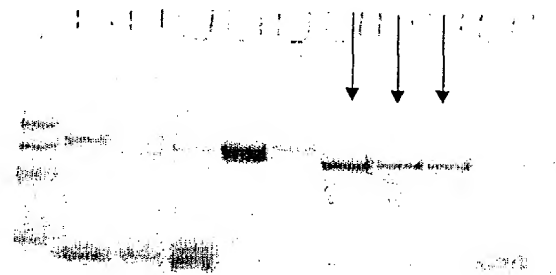


FIG. 123A

KDa P I

115-
84-
62-
51-
38-
26-
20-

FIG. 123B

67.38-GST
67.38-GST

123/169

FIGURE 124

Fig. 124A

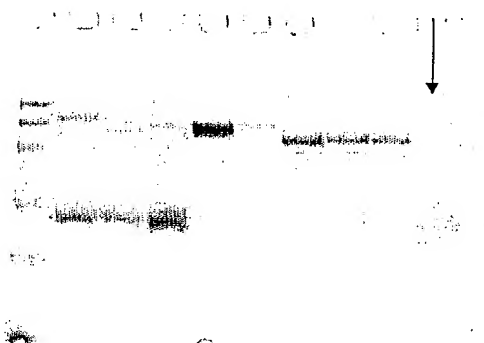
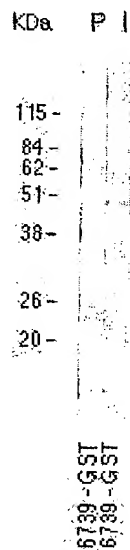


Fig. 124B



124/169

FIGURE 125

Fig. 125A

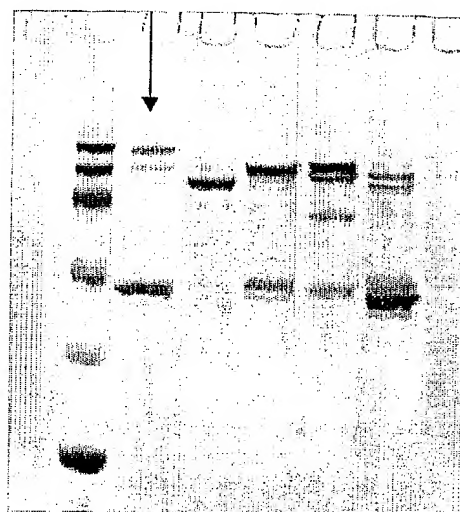
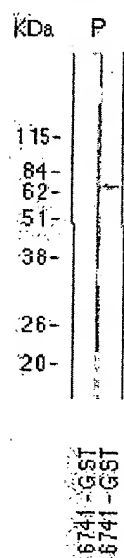


Fig. 125B



125/169

FIGURE 126

Fig. 126A

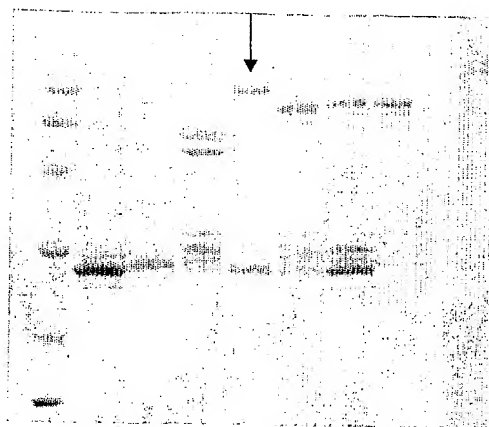
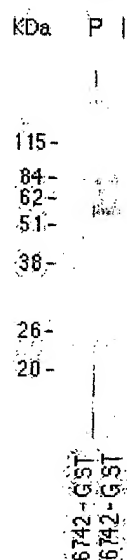


Fig. 126B



126/169

FIGURE 127



FIG. 127A

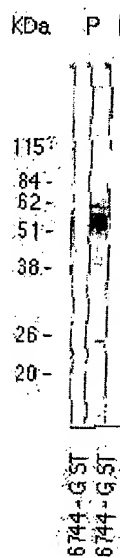


FIG. 127B

127/169

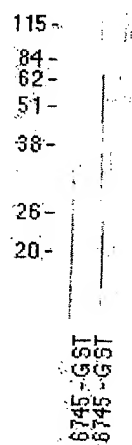
FIGURE 128

Fig. 128A



KDa P I

Fig. 128B



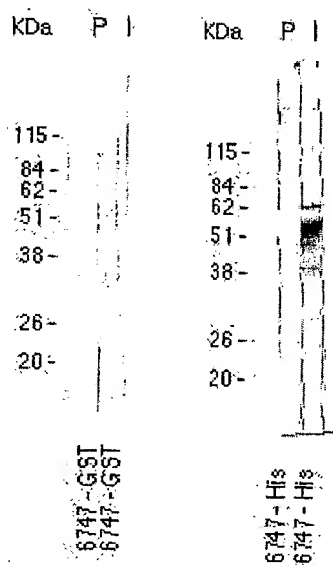
128/169

FIGURE 129

Fig. 129A



Fig. 129B



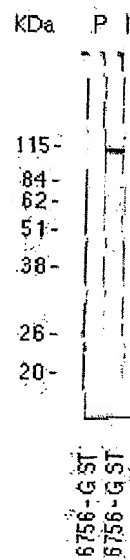
129/169

FIGURE 130

Fig. 130A



Fig. 130B



130/169

FIGURE 131

FIG. 131A

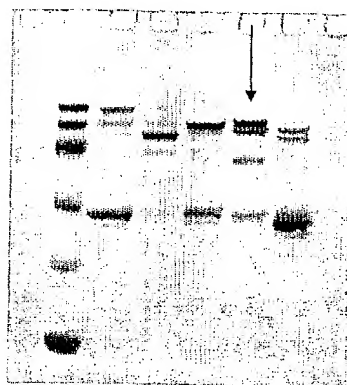
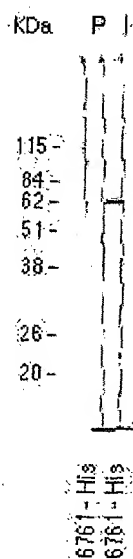


FIG. 131B



131/169

FIGURE 132

Fig. 132A

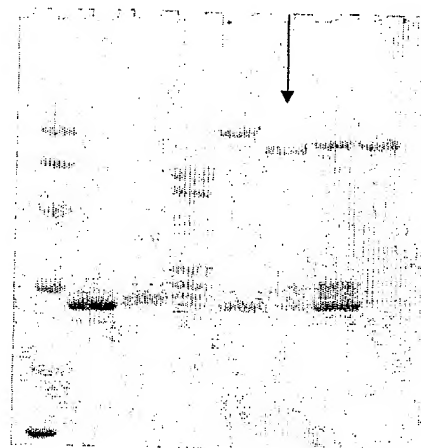
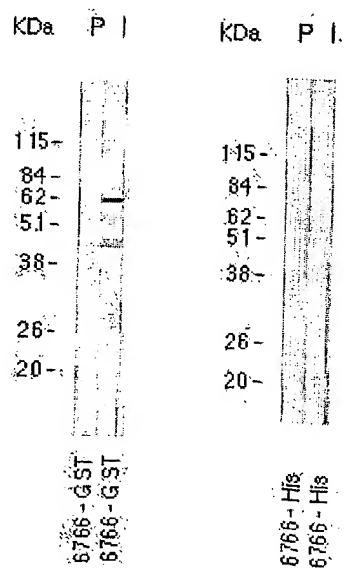


Fig. 132B



132/169

FIGURE 133

Fig. 133A

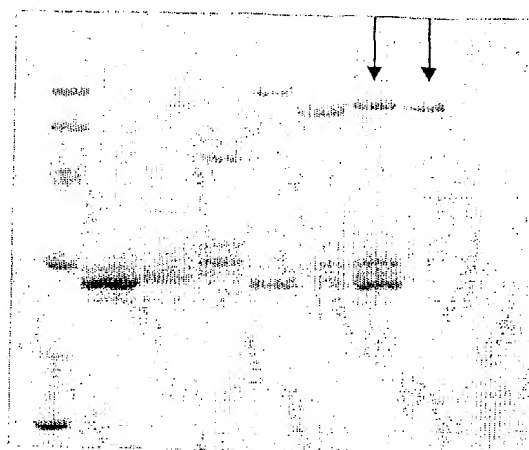
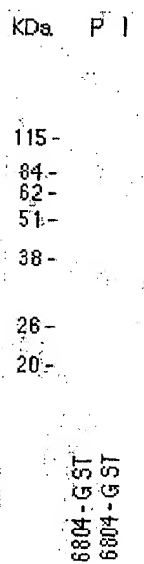


Fig. 133B



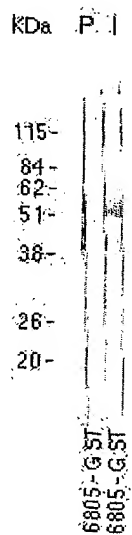
133/169

FIGURE 134

Fig. 134A



Fig. 134B



134/169

FIGURE 135

Fig. 135A

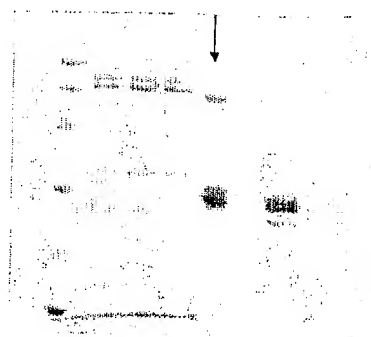
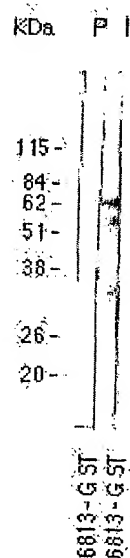


Fig. 135B



135/169

FIGURE 136

Fig. 136A

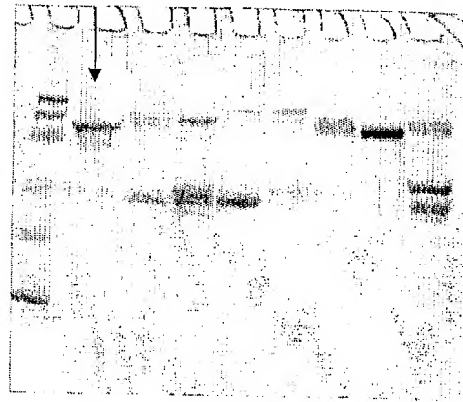
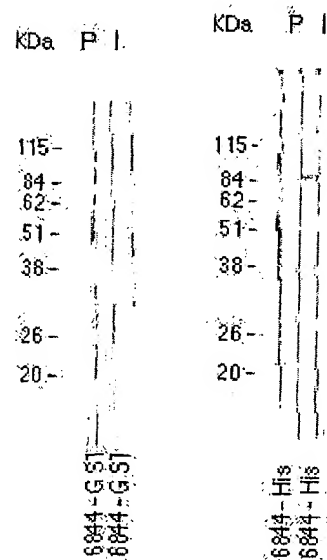


Fig. 136B



136/169

FIGURE 137

FIG. 137A

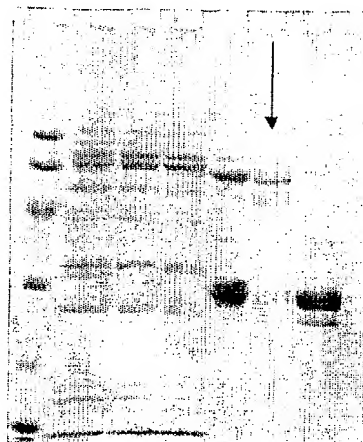
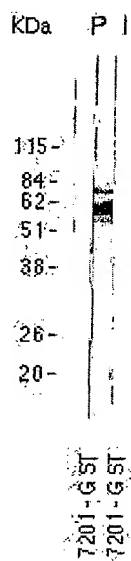


FIG. 137B



137/169

FIGURE 138

Fig. 138A

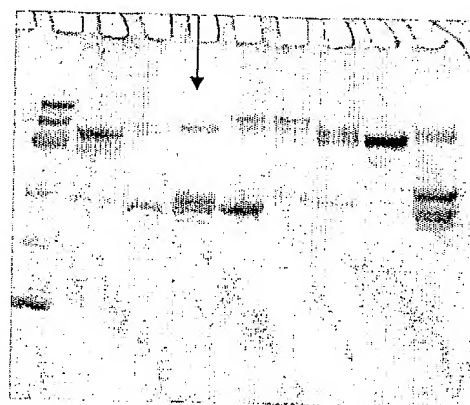
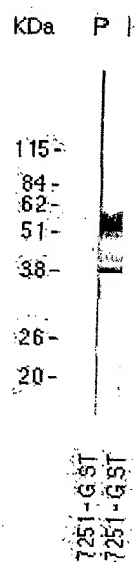


Fig. 138B



138/169

FIGURE 139

Fig. 139A

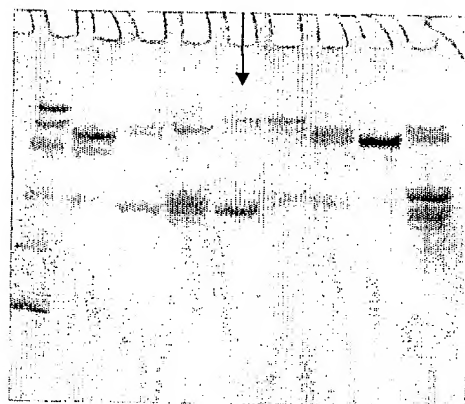
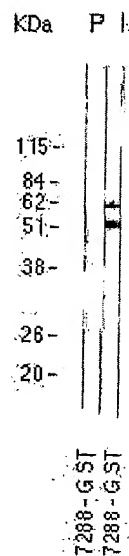


Fig. 139B



139/169

FIGURE 140

Fig. 140A

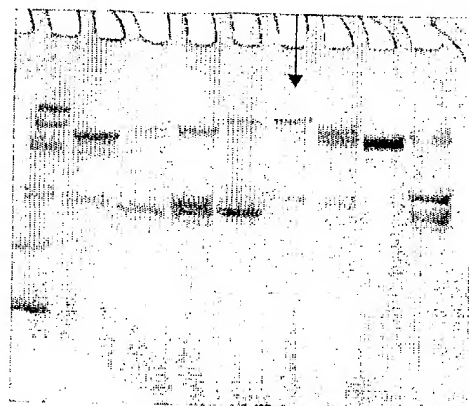
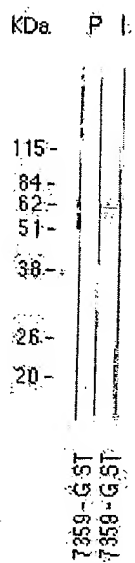


Fig. 140B



140/169

FIGURE 141

Fig. 141A

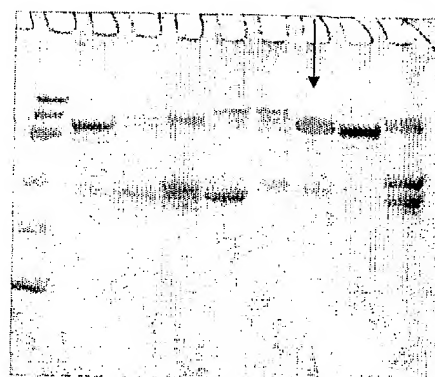
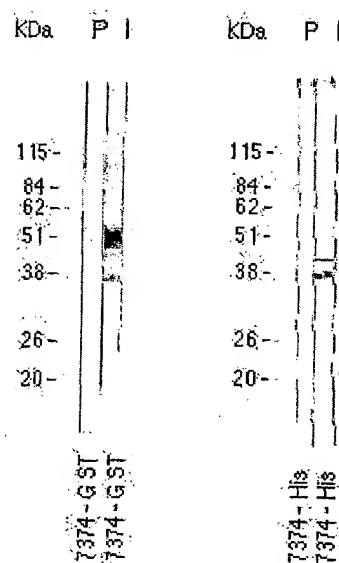


Fig. 141B



141/169

FIGURE 142

FIG. 142A

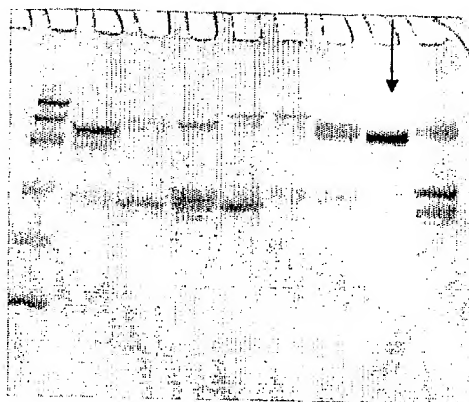
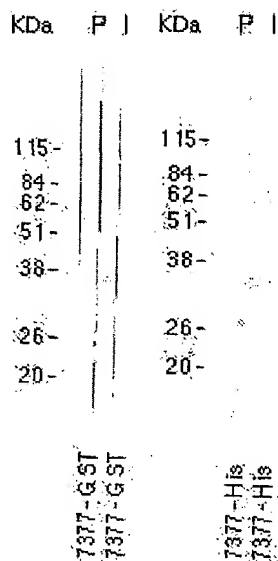


FIG. 142B



142/169

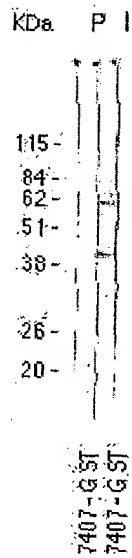
FIGURE 143



Fig. 143A



Fig. 143B



143/169

FIGURE 144

FIG. 144A

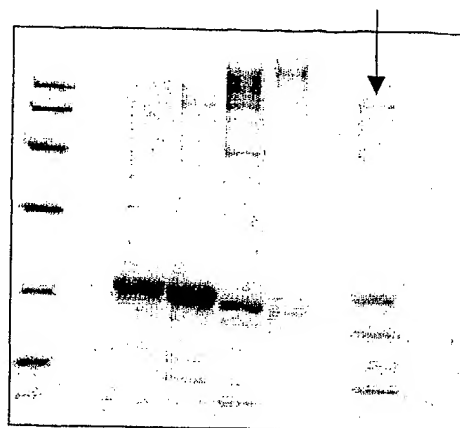
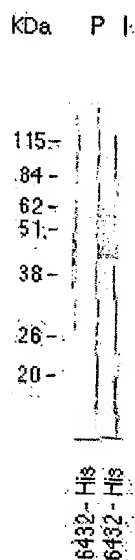


FIG. 144B



144/169

FIGURE 145

FIG. 145A

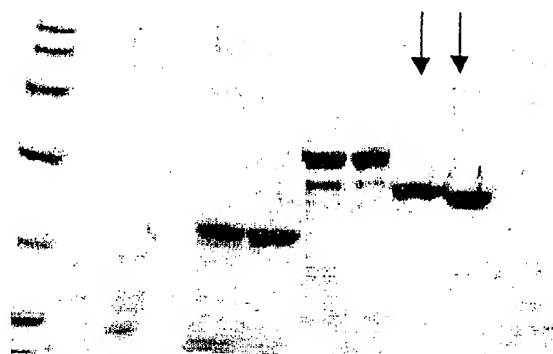
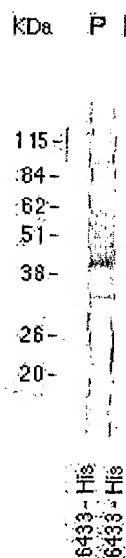


Fig. 145B



145/169

FIGURE 146

FIG. 146A

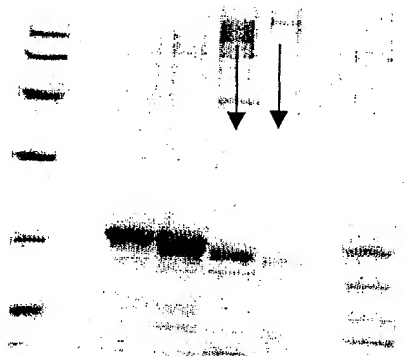
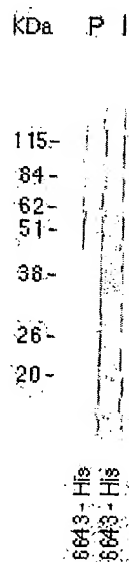


FIG. 146B



146/169

FIGURE 147

FIG. 147A

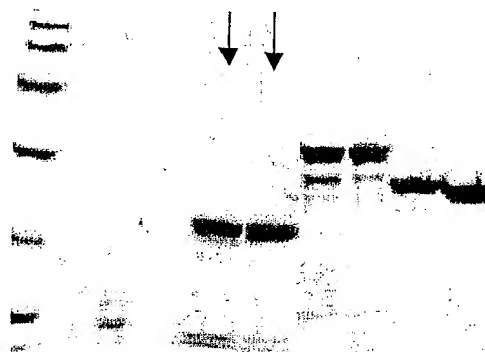
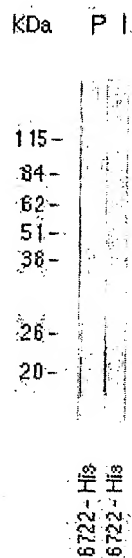


FIG. 147B



147/169

FIGURE 148

Fig. 148A

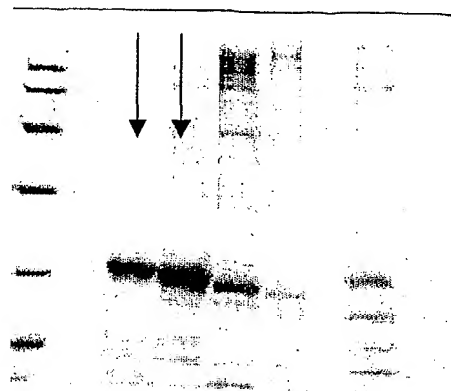
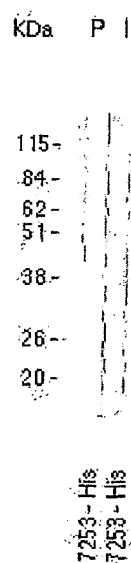


Fig. 148B



148/169

FIGURE 149

FIG. 149A

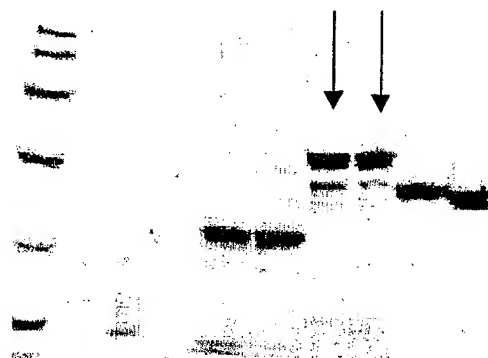
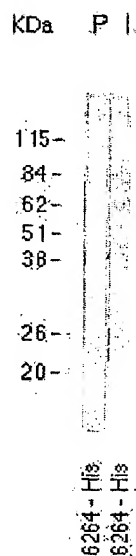


FIG. 149B



149/169

FIGURE 150

FIG. 150A

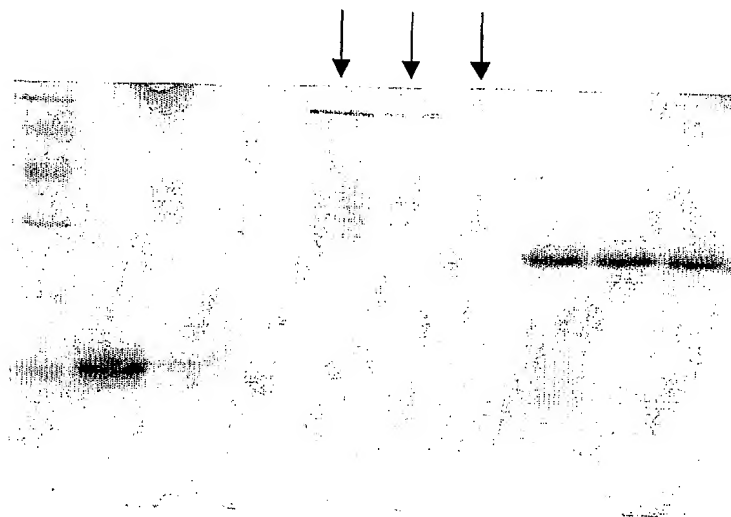
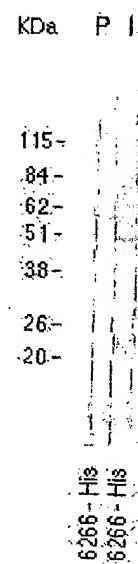


Fig. 150B



150/169

FIGURE 151

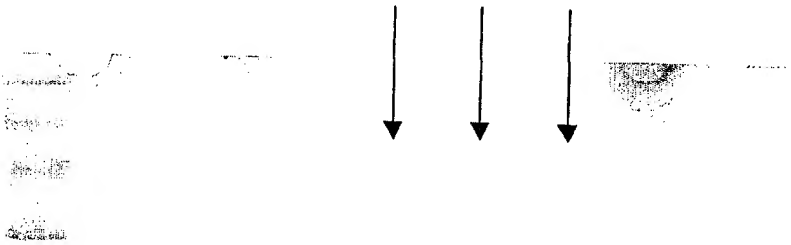


FIG. 151A

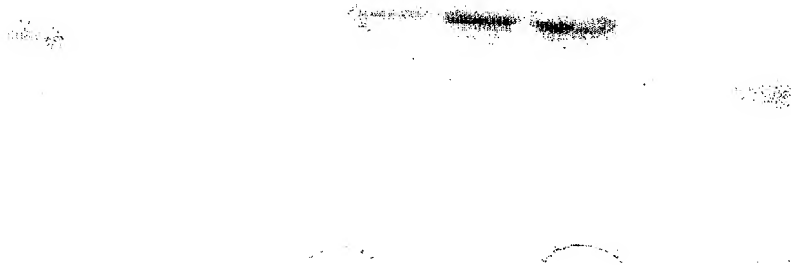
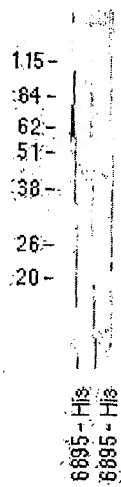


FIG. 151B

KDa P I



151/169

FIGURE 152

Fig. 152A

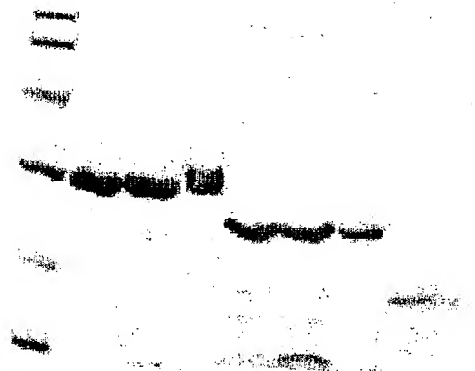


Fig. 152B

kDa P I

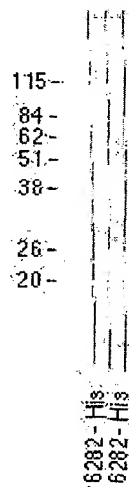
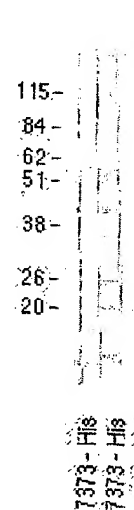


FIGURE 153

kDa P I



152/169

FIGURE 154

Fig. 154A

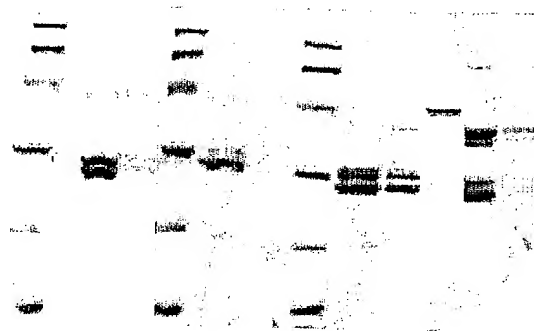


Fig. 154B

KDa P I

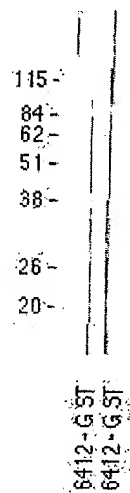


FIGURE 155

KDa P I

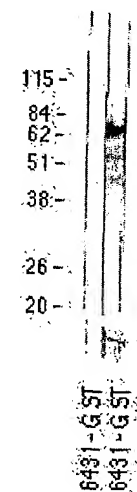


FIGURE 156

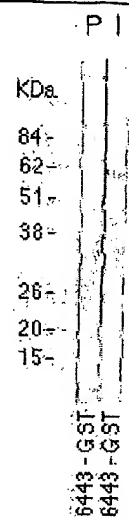


FIGURE 157

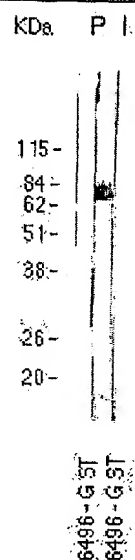
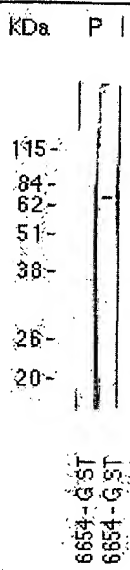


FIGURE 158



154/169

FIGURE 159

Fig. 159A

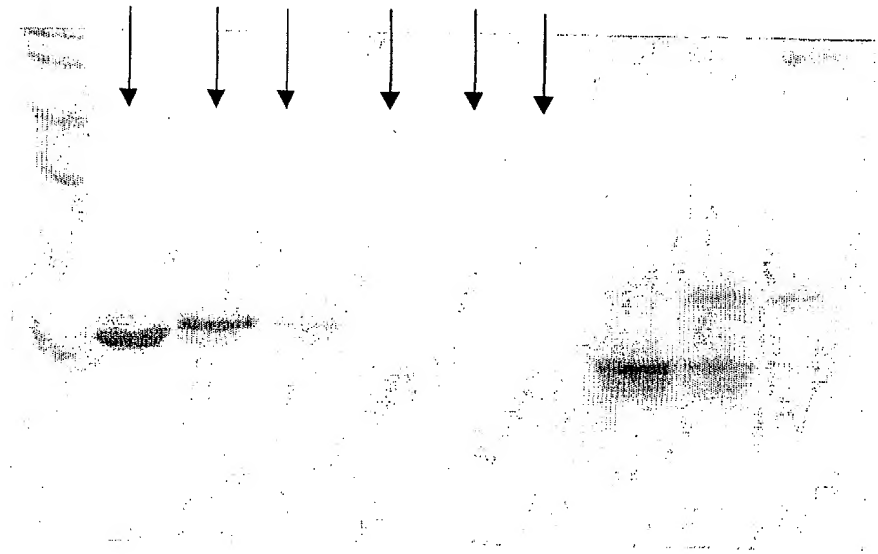


Fig. 159B

KDa P I

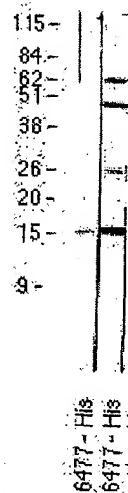
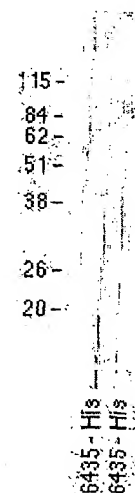


FIGURE 160

KDa P I



155/169

FIGURE 161

Fig. 161A

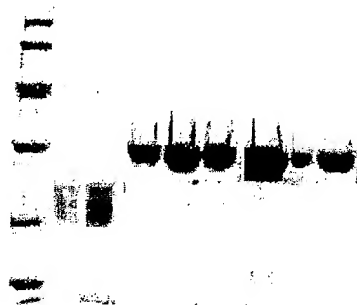


Fig. 161B

kDa P I

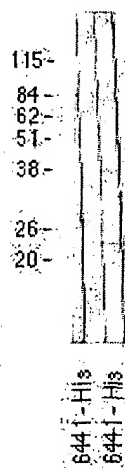


FIGURE 162

kDa P I

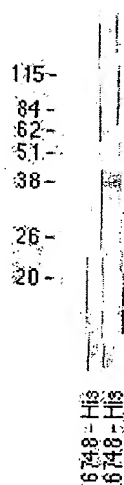
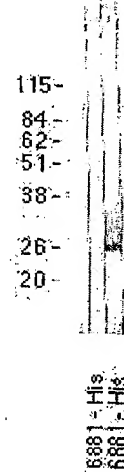


FIGURE 163

kDa P I



156/169

FIGURE 164

Fig. 164A

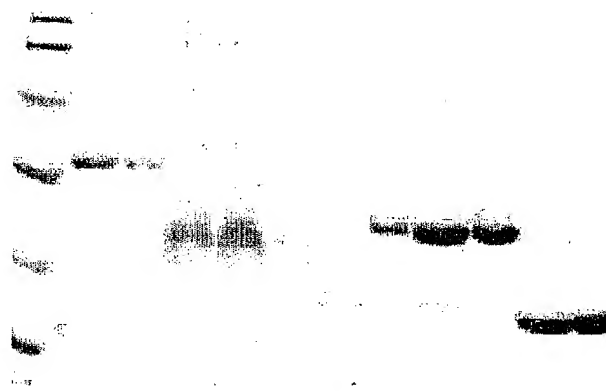


FIG. 164B

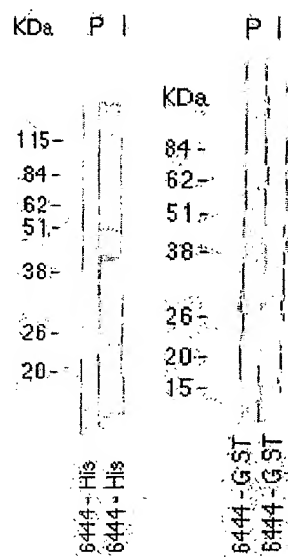


FIGURE 165

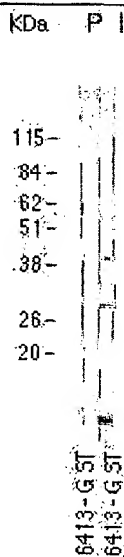
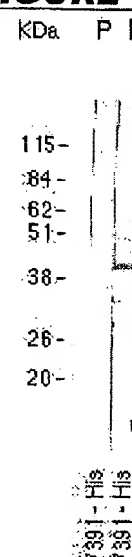


FIGURE 166



157/169

FIGURE 167

Fig. 167A

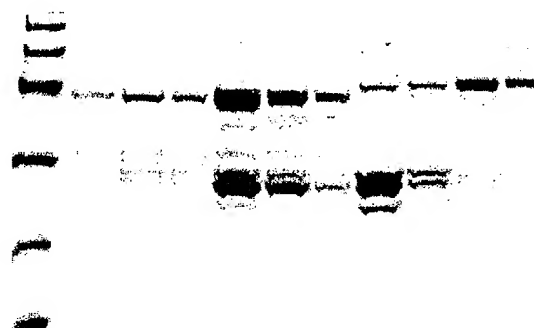


FIG. 167B

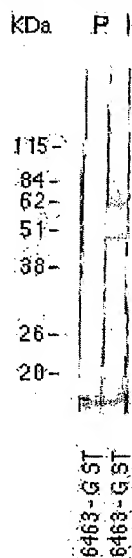
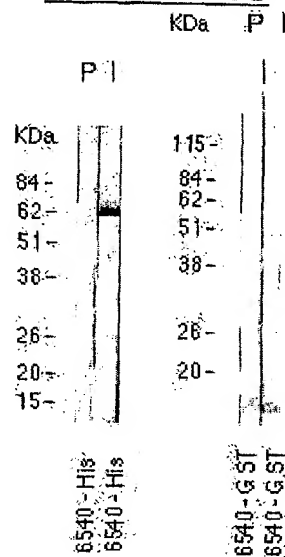


FIGURE 168



158/169

FIGURE 169

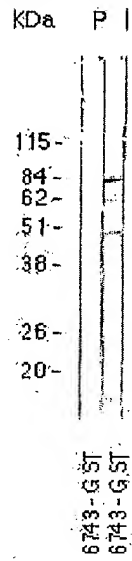
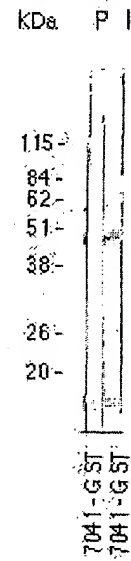


FIGURE 170



159/169

FIGURE 171

Fig. 171A

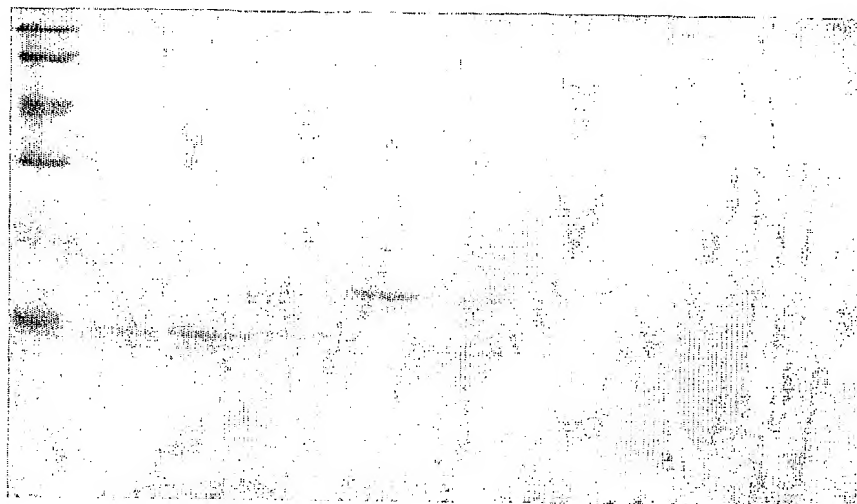


Fig. 171B

KDa P I

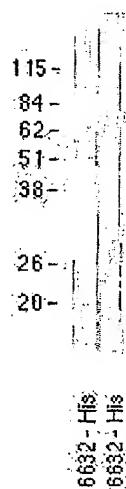


FIGURE 172

KDa P I

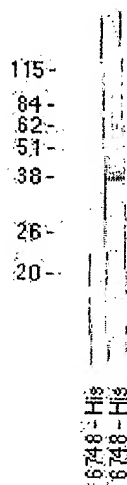
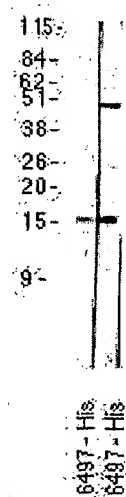


FIGURE 173

KDa P I



160/169

FIGURE 174

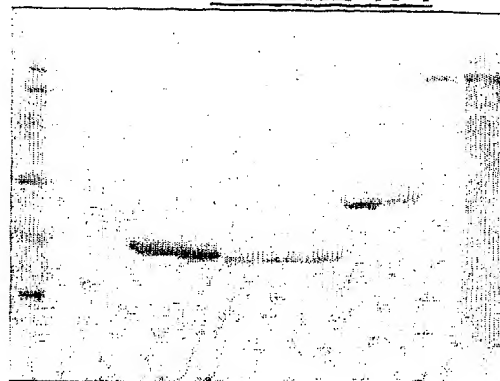


Fig. 174A

FIG. 174B

P I

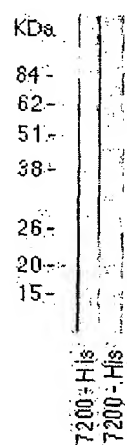


FIGURE 175

P I

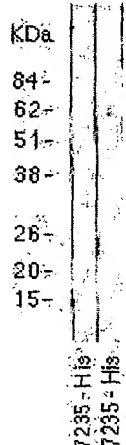


FIGURE 176

P I

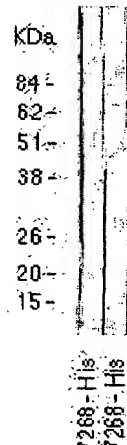


FIGURE 177

P I

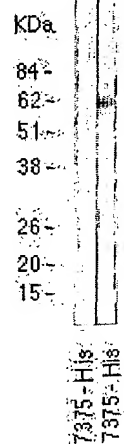
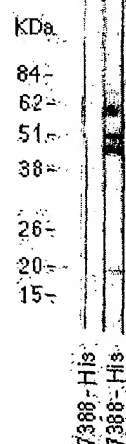


FIGURE 178

P I



161/169

FIGURE 179

Fig. 179A

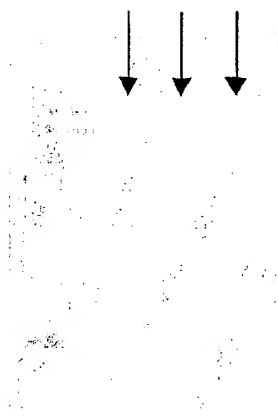
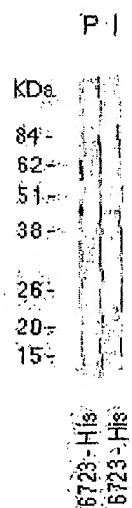


Fig. 179B



162/169

FIGURE 180

Fig. 180A

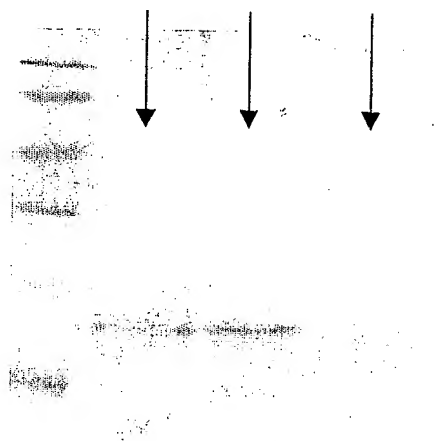
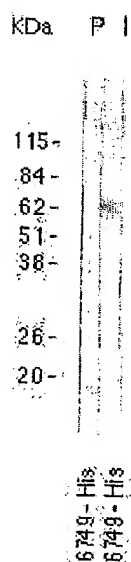


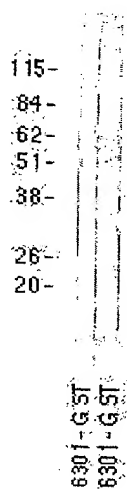
FIG. 180B



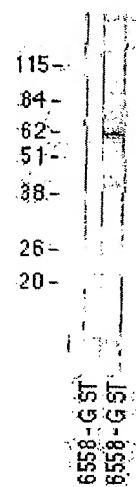
163/169

FIGURE 181

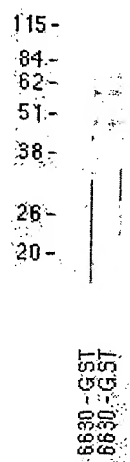
KDa P I

**FIGURE 182**

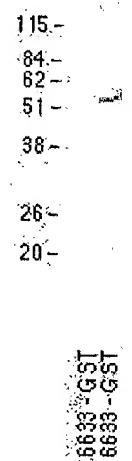
KDa P I

**FIGURE 183**

KDa P I

**FIGURE 184**

KDa P I



WO 02/02606

PCT/IB01/01445

164/169

FIGURE 185

KDa P I

115-

84-

62-

51-

38-

26-

20-

6642-GST
6642-GST

165/169

FIGURE 186

FIG. 186A

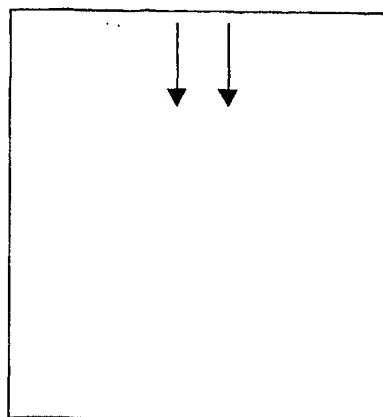
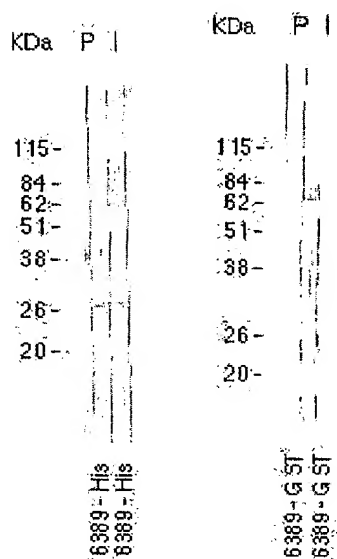


FIG. 186B



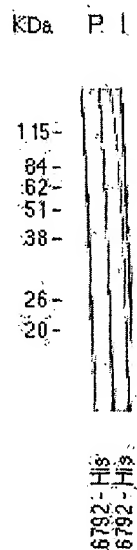
166/169

FIGURE 187

Fig. 187A



Fig. 187B



167/169

FIGURE 188

Fig. 188A

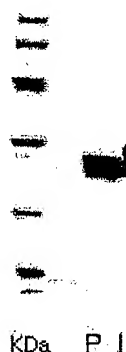
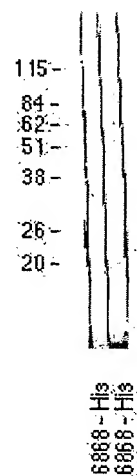


Fig. 188B



168/169

FIGURE 189

FIG. 189A

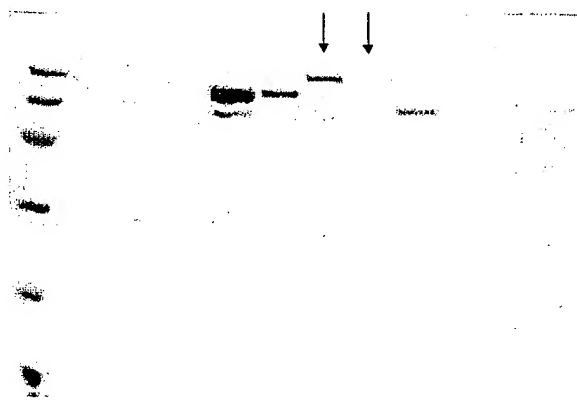
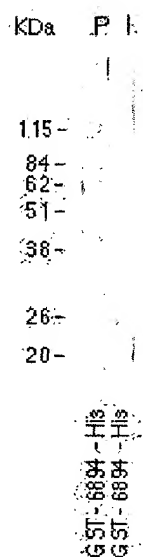
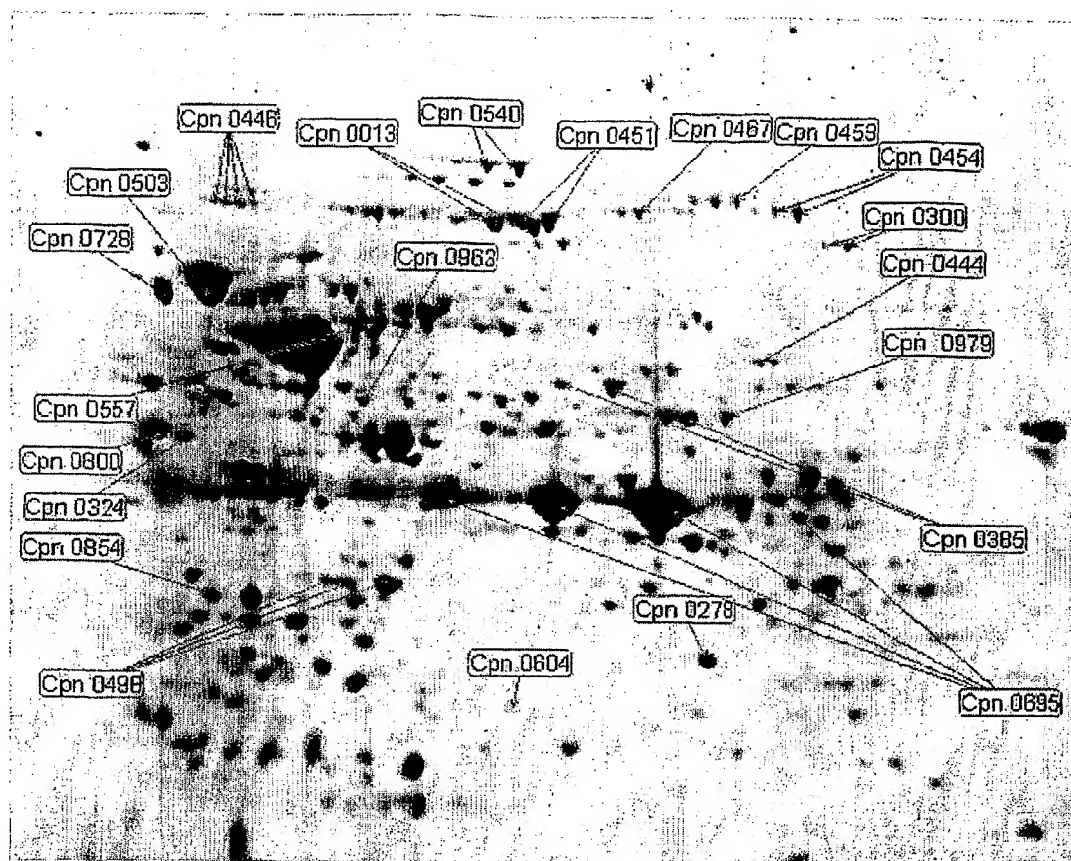


FIG. 189B



169/169

FIGURE 190**FIGURE 191**

```

SVIVG.VSTNSEHRYHAFQYADGQMVDLGLGCGPESYAQGVSGDGK
KVIIVG.HSTRIDGEYRAFKEYVDGRMIDGLGCGSASFAGVSDDGK
KVIIVG.RSETYYGEVHAFCHKNGVMSDGLGCGSYSAAKGVSATGK
KVIIVG.WSTTNNGETHAFMHKDETMHDLGLGCGGFSVATGVSDGCR
TLIVGSMESTITRKTAVKQVNNVPTYLGLGCGDASTGLYISGDGT

```

